UNIVERSITY OF CAGLIARI
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
PH.D. COURSE IN COMPUTER SCIENCE
CYCLE XXVIII

PH.D. THESIS

# CLOUD-BASED SOLUTIONS IMPROVING TRANSPARENCY, OPENNESS AND EFFICIENCY OF OPEN GOVERNMENT DATA

CANDIDATE
Gianfranco Garau

PH.D. SUPERVISOR
Prof. Nicoletta Dessì

PH.D. COORDINATOR
Prof. G. Michele Pinna

S.S.D. INF/01
2015 – 2016

# Abstract

A central pillar of open government programs is the disclosure of data held by public agencies using Information and Communication Technologies (ICT). This disclosure relies on the creation of open data portals (e.g. Data.gov) and has subsequently been associated with the expression Open Government Data (OGD). The overall goal of these governmental initiatives is not limited to enhance transparency of public sectors but aims to raise awareness of how released data can be put to use in order to enable the creation of new products and services by private sectors.

Despite the usage of technological platforms to facilitate access to government data, open data portals continue to be organized in order to serve the goals of public agencies without opening the doors to public accountability, information transparency, public scrutiny, etc. This thesis considers the basic aspects of OGD including the definition of technical models for organizing such complex contexts, the identification of techniques for combining data from several portals and the proposal of user interfaces that focus on citizen-centred usability.

In order to deal with the above issues, this thesis presents a holistic approach to OGD that aims to go beyond problems inherent their simple disclosure by providing a tentative answer to the following questions:

1) To what extent do the OGD-based applications contribute towards the creation of innovative, value-added services?

2) What technical solutions could increase the strength of this contribution?

3) Can Web 2.0 and Cloud technologies favour the development of OGD apps?

4) How should be designed a common framework for developing OGD apps that rely on multiple OGD portals and external web resources?

In particular, this thesis is focused on devising computational environments that leverage the content of OGD portals (supporting the initial phase of data disclosure) for the creation of new services that add value to the original data.

The thesis is organized as follows. In order to offer a general view about OGD, some important aspects about open data initiatives are presented including their state of art, the existing approaches for publishing and consuming OGD across web resources, and the factors shaping the value generated through government data portals.

Then, an architectural framework is proposed that gathers OGD from multiple sites and supports the development of cloud-based apps that leverage these data according to potentially different exploitation roots ranging from traditional business to specialized supports for citizens.

The proposed framework is validated by two cloud-based apps, namely ODMap (Open Data Mapping) and NESSIE (A Network-based Environment Supporting Spatial Information's Exploration). In particular, ODMap supports citizens in searching and accessing OGD from several web sites. NESSIE organizes data captured from real estate agencies and public agencies (i.e. municipalities, cadastral offices and chambers of commerce) in order to provide citizens with a geographic representation of real estate offers and relevant statistics about the price trend.

# Contents

# List of Figures

# List of Tables

# Acronym

| | |
|---|---|
| AJAX | Asynchronous JavaScript and XML |
| API | Application Programming Interface |
| DCAT | Data Catalog Vocabulary |
| FOI | Freedom of Information |
| GAE | Google App Engine |
| IaaS | Infrastructure as a Service |
| JSON | JavaScript Object Notation |
| NoSQL | Not Only Structured Query Language (i.e., non-relational) |
| ORM | Object-Relational Mapping |
| PaaS | Platform as a Service |
| RDF | Resource Description Framework |
| REST | REpresentational State Transfer |
| RPC | Remote Procedure Call |
| SaaS | Software as a Service |
| URI | Uniform Resource Identifier |

# Introduction

The term "open government" dates back to the initiative of U.S. President Obama having announced a new era of participation, transparency, and collaboration in 2009 [1].

Since then, many states are recognizing the importance of involving citizens and external parties in governmental processes and experimenting new approaches for the success of e-government initiatives. The concept of open government entails fundamental changes in the public service however most initiatives have been restricted to the launch of Open Data Portals [2] that present government data to the public. A central pillar of these initiatives is the active disclosure of data held by public agencies allowing the general public to monitor their actions and performances [3][4]. The content of Open Data Portals has subsequently become associated with the expression Open Government Data (OGD) as these web resources provide access to data released by various public sector authorities. Widely adopted, open data portals are now visible with various governments and an increasing number of countries are announcing new initiatives towards opening up their public information. Open Data Portals aim to introduce a shift from the static information provision towards the delivery of citizen-centric services to meet public-sector objectives such as efficiency, transparency and accountability [5].

Beside the OGD should be used for better understanding what the government does and how well it performs, the overall goal of these governmental initiatives is not limited to enhance transparency of public sectors but it aims to raise awareness of how released data can be put to use. Indeed, data collected by governments represents a particularly significant resource when used in conjunction with information from other sources. Following the concept of open data which anyone is free to use, the published OGD could be accessed and reused for any purpose by everyone interested in harnessing OGD for the creation of new products and services which gather information from several Open Data Portals and web sources. According to this point of view, OGD is also considered as an important source of economic enabling new forms of entrepreneurships and social innovation.

To summarize and state again the main points of the above discourse, OGD might serve the purposes of increasing [6]:

(i) Transparency - Citizens must be allowed to access, share and reuse government data. The active disclosure of data held by public agencies should enable citizens to monitor their actions and performances [4][7].

(ii) Commercial and social value - Data produced and gathered by governments span several domains including education, transports, territory, budget, etc. Having a social and economic value, this data can be used as a basis for the creation of new services that add value to the original data and can be commercialized. The reuse of OGD has a great potential for a number of different commercial and social activities. As well, it fosters the creation of innovative business and social interactions that promote the creation of innovative value added-services by stakeholders [8].

(iii) Participatory Governance - The publication of OGD gives citizens the opportunity to be involved in government decision-making processes. Open data initiatives, such as portals contribute to improve the decision making of both governments and empower citizens, which are expected to give insights into how to improve government performance [9].

Despite their apparent success, Open Data Portals have been criticized [10] because their content continues to serve the existing goals of public agencies whose common assumption is that the disclosure of the government data will lead to increase the transparency and the public accountability. In the last few years, such assumption contributed to reduce the initial enthusiasm behind the open government. Restricted to tackling the ability of the governments to present their data to the public, these portals are data silos offering a lot of datasets of a summary nature and pertaining to a plethora of activities.

With the aim of promoting data search, these datasets have been enriched with metadata standards such as Dublin Core [11]. However, the flatness of metadata and the absence of integrity checking mechanisms make quite impossible the search of information contained in different datasets in which data is stored in different formats.

In addition to persistent obstacles including the limited budget of government agencies and the lack of standard frameworks for publishing data, a major technical barrier that prevents the exploitation of OGD remains their heterogeneity. Overcoming this heterogeneity represents a challenge because it requires to devise automatic mechanisms for capturing information from OGD and new practices for improving the usability of public datasets whose search is often unwieldy. Ideally, OGD from several portals would be integrated in a comprehensive database with a nicely integrated schema that covers all metadata of available documents. However, the implementation of this solution is challenging because 1) documents and associated metadata change frequently, 2) metadata standards provide limited information, 3) it is very difficult to assure integrity of the associated documents using integrity checking mechanisms. Many initiatives only result in implementing Open Data Portals (often little more than websites) which contain miscellaneous and heterogeneous data files. No attention is paid to very important aspects of opening governmental data such as the definition of technical models for organizing such complex contexts, the identification of techniques for combining data from several portals, the proposal of user interfaces that focus on the citizen-centred usability [12][13].

This thesis aims to present a holistic approach for dealing with the above aspects that encompass the balance of various components including technical solutions, political and legal factors, economic and financial aspects and citizen interests. In particular the thesis focuses on the following questions:

1) To what extent do the OGD-based applications (namely OGD apps from now on) contribute to the creation of innovative, value-added services and what technical solutions could strengthen this contribution?

2) How can Web 2.0 and Cloud technologies favour the development of OGD apps?

3) How would a common framework for developing OGD apps that rely on multiple OGD portals and external web resources?

Towards facing the above questions, the thesis tries to improve understanding of what restricts the effects of open data initiatives by discussing different aspects of such initiatives such as their state of art, the existing approaches for publishing and consuming OGD across web resources and the factors shaping the value generated through government data portals. The aim is to present a perspective that broadens and sharpens the basic knowledge about the nature of OGD and the above mentioned technical barriers.

To demonstrate the potential resulting from new technologies and new practices in guiding the design of innovative OGD apps, the thesis presents an architectural framework that harnesses OGD both structurally and operationally in order to potentially support the development of cloud-based OGD apps that follow different exploitation roots ranging from traditional business to citizen supports. The prerequisites for experimenting cloud technologies are 1) the government's needs to rely on ICT advances for promoting open government and 2) the necessity of integrating OGD through web applications that enhance the quality of the relationships between citizens and their governments.

The proposed framework is based on the following strengths:

1)the creation of a catalogue that serves as a point of convergence for metadata harvested from government portals and public resources;

2)the full exploitation of cloud technology for the development and the deployment of OGD apps.

3) The use of the NoSQL database offered by the cloud environment to achieve great flexibility in organizing the above mentioned catalogue while avoiding the definition of an a priori metadata schema.

Two case studies are presented that validate the proposed approach. The first case studies illustrates ODMap (Open Data Mapping), a prototypical cloud-based application that supports citizens in searching and accessing OGD stored in several web sites. The second case study presents NESSIE (Network-based Environment Supporting Spatial Information's Exploration) a cloud-based application that captures data from real estate agencies and OGD data from municipalities, cadastral offices and chambers of commerce. NESSIE provides the user is provided with a geographic representation of real estate offers and relevant statistics about the price trends.

ODMap and NESSIE exhibit spatial features in order promote a better understanding of delivered information. Users are allowed to visualize maps, define and store geographical zones as well as to formulate spatial queries. Both applications have been developed and deployed on top of Google App Engine (GAE) [14], a cloud computing environment which provides a platform-as-a-service for developing and hosting web applications.

The thesis is organized as follows. Chapter 1 presents the basic characteristics of OGD initiatives including its main phases and parties involved and the technical aspects that are relevant in preventing the full application of open data paradigm. Chapter 2 presents the most popular platforms for publishing data and some value-added applications that rely on OGD for creating services of interest for the citizens. A background on Cloud Technologies is introduced in chapter 3 that also motivates the introduction of such technologies for improving the functionality of open government portals. At the end of the chapter 3, a framework is proposed that harnesses an integrated set of cloud-based services to lower the complexity of searching and integrating OGD regardless of its format and location. The framework aims to facilitate the design of OGD-based apps that support query and navigation with high level of flexibility to the user needs. Towards a validation of the proposed framework, chapter 4 presents a case study that introduces ODMap (Open Data Mapping), a prototypical cloud-based application featured by dynamic patterns of interaction to increase the interest of the users in gathering data from multiple government portals. A further case study is presented in chapter 5 that describes NESSIE, a cloud-based application for supporting citizens and stakeholders in the evaluation of real estate offers. Finally, the conclusions are presented in chapter 6.

# 1 Background

OGD is a relatively new concept. It dates back to 1966 the first attempts to guarantee public access to government information under the U.S. Federal government's Freedom of Information Act (FOIA) [15]. Specifically, interested citizens were in charge of requiring information by compiling specific files and sending it to the administration that owned the data. Most of the received requests were for unstructured data such as scanned documents and images about the car fines by police. In 2003 the Public Sector Information Directive provided a common legal framework to support a European market for government-held data [16]. Starting in 2006 and according with the ICT improvements, many public administrations began making available datasets on web resources. As structured data can be published easily, a considerable amount of data was made available in tabular format or in a spreadsheet. In 2007, thirty open government advocates gathered in California and wrote a set of eight principles on open government data with the aim of encouraging a more robust understanding of why open government data is essential to democracy [17].

A number of open data government actions have been performed in recent years for enhancing reuse and transparency of government data. Some actions focus only on their impact in the government and public life without attention to the quality of the disclosed content. In turn, it is possible to observe narrowly focused initiatives that result in the implementation of irrelevant websites and/or heterogeneous data files without attention for the political and the social outcomes of the use of the published data. Supporting OGD initiatives through the use of ICT it requires a good understanding of how the things work in order to balance political and technical aspects.

This section aims to contribute to increasing this understanding by presenting the basic characteristics of the common OGD process adopted from open government initiatives, existing research work and standards for practice.

## 1.1 Open Government Initiatives

As previous mentioned, a new era of participation, transparency, and collaboration was promoted in 2009 by the "Memorandum on transparency and Open Government" signed by the US President OBAMA on his first day in office. Notably, the Obama administration in May of the same year, lunched the data.gov website hosting more than 186.000 datasets [18]. The last five years have seen a remarkable development in the disclosure of government data. The availability of open data has grown significantly and increasing data openness is now a political priority in the US [1], UK [19] and EU [20] and many other states. Some motivations are that this policy pushes the transparency of public administrations and speeds up the processes of innovation and market expansion through the proactive release of government data. In 2011 the Open Government Partnership was launched to provide an international platform for domestic reformers and make their governments more open, accountable and responsive to citizens [21]. Similarly, in 2013 the Open Data Charter was signed by G8 leaders to promote transparency, innovation and accountability [22]. The Directive on the re-use of public sector information (PSI Directive) has been adopted EU in 2003 [16] and revised in 2013 [20]. The aim was to introduce a common legislative framework regulating how public sector bodies should make their information available and remove barriers through their re-use. All EU member states have implemented the PSI Directive with their national legal orders. It is based on two fundamental points of internal market: transparency and fair competition.

Public data originate from public administration in many different domains [23] including but not limited to data about the following domains:

- Culture: Data about cultural works and artefacts - for example titles and authors - generally collected and held by galleries, libraries, archives and museums.
- Science: Data that is produced as part of scientific research from astronomy to zoology.
- Finance: Data such as government accounts (expenditure and revenue) and information on financial markets (stocks, shares, bonds, etc.).
- Statistics: Data produced by statistical offices such as the census and key socioeconomic indicators.
- Weather: The many types of information used to understand and predict the weather and climate.
- Environment: Information related to the natural environment such presence and level of pollutants, the quality and rivers and seas.
- Transport: Data such as timetables, routes and on-time statistics.

Table 1 shows the dataset categories in highest demand across EU from [24]. In 2009, the previous mentioned directive of the president Obama has constrained governments to take a different approach and face the significant challenge of publishing their data in open data portals. In 2009, the previous mentioned directive of the president Obama has constrained governments to take a different approach and face the significant challenge of publishing their data in open data portals. This challenge results in converting the thousand of datasets typically owned by a government office into data that must conform to the specifications of "open data" showed in Table 2 i.e. the 5-star open data scheme suggested by Tim Berners-Lee's [25] in 2010.

| Category | Examples of datasets |
|---|---|
| 1. Geospatial data | Postcodes, national and local maps (cadastral, topographic, marine, administrative boundaries, etc.) |
| 2 Earth observation and Environment | Space and in situ data (monitoring of weather, land and water quality, energy consumption, emission levels, etc.) |
| 3. Transport data | Public transport timetables (all modes of transport) at national, regional and local levels, road works, traffic information, etc. |
| 4. Statistics | National, regional and local statistical data with main demographic and economic indicators (GDP, age, health, unemployment, income, education, etc.) |
| 5. Companies | Company and business registers (lists of registered companies, ownership and management data, registration identifiers, balance sheets, etc. |

**Table 1 - The dataset categories in highest demand across EU.**

| | | |
|---|---|---|
| ★ | OL | Open Licence - Open data with an open license published on the web in whatever format. |
| ★★ | OL MR | Machine Readable - Data presented in a machine-readable format (e.g. excel instead of an image). |
| ★★★ | OL MR OF | Open Format - Same as two stars but the data must be in a non-proprietary format (e.g. CSV instead of excel). |
| ★★★★ | OL MR OF URI | Uniform Resource Identifier - All above but the data must be marked with URIs so that it can be pointed at. |
| ★★★★★ | OL MR OF URI LD | Linked Data - All above but the data must be linked to other data to provide context. |

**Table 2 - Tim Berners-Lee's 5-star Open Data Scheme.**

Given the special nature of government data, the Working Group on Open Government Data [17] established the unique intersection of open government and open data e.g. the 8 principles showed in Table 3.

| | | |
|---|---|---|
| 1 | **Complete** | All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations. |
| 2 | **Primary** | Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms. |
| 3 | **Timely** | Data is made available as quickly as necessary to preserve the value of the data. |
| 4 | **Accessible** | Data is available to the widest range of users for the widest range of purposes. |
| 5 | **Machine processable** | Data is reasonably structured to allow automated processing. |
| 6 | **Non-discriminatory** | Data is available to anyone, with no requirement of registration. |
| 7 | **Non-proprietary** | Data is available in a format over which no entity has exclusive control. |
| 8 | **License-free** | Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed. |

**Table 3 - The eight principles of Open Government Data.**

Moreover, the Working Group has formulated the additional seven principles showed in Table 4.

| | | |
|---|---|---|
| a | **Online & free** | Information is not meaningfully public if it is not available on the Internet at no charge, or at least no more than the marginal cost of reproduction. It should also be findable. |
| b | **Permanent** | Data should be made available at a stable Internet location indefinitely and in a stable data format for as long as possible. |
| c | **Trusted** | Digital signatures help the public validate the source of the data they find so that they can trust that the data has not been modified since it was published. |
| d | **A Presumption of Openness** | The presumption of openness rests on laws like the Freedom of Information Act (FOIA), procedures including records management, and tools such as data catalogues. |
| e | **Documented** | Documentation about the format and meaning of data goes a long way to making the data useful. |
| f | **Safe to Open** | Data online should be always published using data formats that do not include executable content. |
| g | **Designed with Public Input** | The public is able to determine which information technologies will be best suited for the applications the public intends to create for itself. |

**Table 4 - Seven additional principles of Open Government Data**

As mentioned in [26] “In order to provide official guidelines, the W3C eGov Interest Group has also developed the following set of steps for publishing open government data, which emphasizes standards and methodologies, with the aim of enabling easier data use by the public:

1. *Identify* - The use of permanent, patterned and/or discoverable URI/URLs enables processes and people to find and consume the data more easily.

2. *Document* - Documentation helps the data to be more understandable and less ambiguous, as well it enables easier data discovery. The use of formats such as XML/RDF would be self-documenting.

3. *Link* - Linked data contains links to other data and documentation, providing context.

4. *Preserve* - The use of versioning of datasets enables data consumers to cite and link to present and past versions, where new and upgraded datasets can refer back to original datasets. Versioning also allows the documentation of changes between versions.

5. *Expose interfaces* – To be discovered and explored easy, published data should be both human-readable and machine-readable. Preferably, data should be published separately from the interface, and external parties should have direct access to raw data. This enables them to build their own interfaces if needed.

6. *Create standard names/URIs for all government objects* - the use of a unique identifier for each object is as important as having information about the object itself. This aids in discoverability, improves metadata, and ensures authenticity.

Among the important initiatives that have as their stated goal the pursuit of innovation through the OGD, it has to be mentioned the Open Data Institute (ODI) [27]. Based in London and founded in 2012, it is an independent, non-partisan and non-profit company that has as co-founders the inventor of the web Tim Berners-Lee and AI expert Sir Nigel Shadbolt. Relying on high profile board members, the ODI is open to individuals, companies and public and private organizations of all sizes that are interested in the development and progress through the data. The ODI aims to bring together companies, organizations and governments on specific areas to address current global challenges. The major contribution is to identify and understand how the data, available or potentially available on the web, may impact the objectives of companies and organizations. To achieve its aims, the ODI defined the concept of data infrastructure as a tool capable to offer open innovation at web-scale. Here the concept of open data is expanded: the data infrastructure, in addition to the data patrimony, includes organizations that operate on open data, manage them and define the guidelines for their use and management. According to the above expansion, the data infrastructure includes technology, processes and organization and can be at the city, regional, national or global level. Moreover, data are seen as a physical infrastructure, such as roads: as well as roads allow people to reach a destination, data allow people to make a decision.

Through its activities, studies and publications, the ODI activity aims at clarifying the terminology around the open data, also preventing possible confusion than may have a negative impact on the dissemination of open data culture. For example, it is important to be aware that the spectrum of data (Figure 1) covers different aspects that could prevent the fully diffusion of the open data culture.

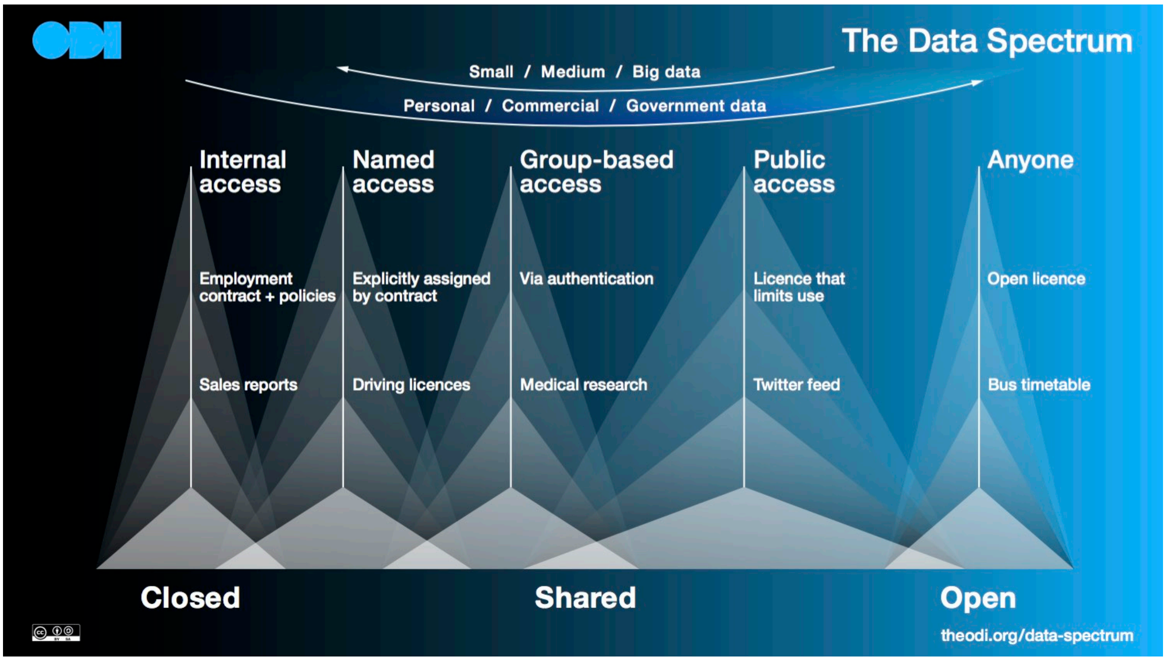



Figure 1 - The ODI Data Spectrum.

This means that not all the data are expected to become open, or eventually, they can become so only after some processes including, for example, data transformations to avoid problems to personal privacy or to corporate strategies.

Another known organization, based in US, is the Sunlight Foundation [28]. This is a national, no-partisan, no-profit organization that aims to make US government and politics more accountable and transparent using open data, technology, policy analysis and journalism. Sunlight Foundation has started his work in 2006 with only focus on the U.S. Congress; his open government work now takes place at the local, state, federal and international levels. The foundation's primary focus is the role of money in politics, but additional critical issues related to open data are also addressed. A notable result of the Sunlight Foundation work is a document about the OD policy guidelines [29] that detail what data should be public, how to make them public and how to implement a successful open data policy.

Currently, several administrations store their data in platforms that capture the legacy data from public sector entities, support gathering and storing data from the government activities and organize datasets to enable easy downloading and sharing by users. Although the number of public institutions and related initiatives to public their data is drastically increasing, it is still a major challenge to fully benefit from open government data and support all the interested stakeholders in accessing and using this data.

PublicData.eu [30] and the recent European Data Portal [31] are Pan-European data portals, which provide access to open, freely reusable datasets from local, regional and national public bodies across Europe.

The Open Data City Census by the Knowledge Foundation contains more than one thousand datasets from more than one hundred cities and represents a good example of how showing the kind of data and the related quality in the cities [32].

Technical, policy, legal, economic, financial, organization and cultural issues constitute a barrier [12][13] as well as the heterogeneous nature of data formats used by public administrations (such as PDF, CSV, spreadsheets, XML and database records). Systems and tools used by governments constitute a further obstacle hindering society from realizing government data transparency. Government platforms appear as large silos of documents with heterogeneous and often unstructured data, in which the information provided is mainly through PDF, HTML, Excel spreadsheets or other print-like format. Much of the presented material is rather general. Monolithic and rigid user interfaces fail to provide support for exploring content or integrating the provided datasets with those provided by other administrations which would allow citizens the effective use of transparency. This motivates why citizens seem to distrust about the effectiveness of the open data initiatives. In this regard, Figure 2, Figure 3 and Figure 4 show the results of a recent survey [33] conducted on 3212 USA citizens.

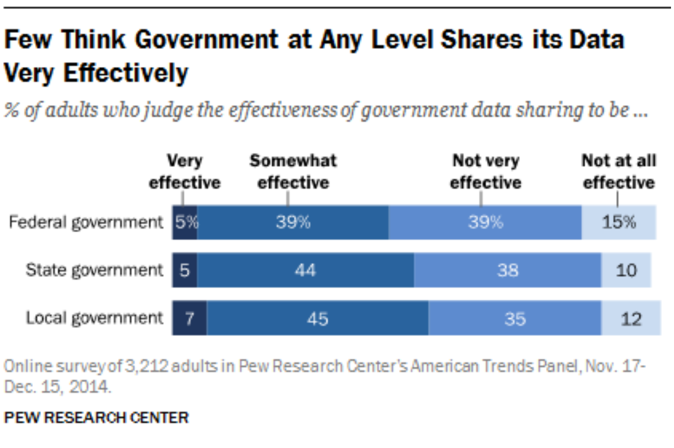



Figure 2 - PEW Survey: Effectiveness of OGD sharing.

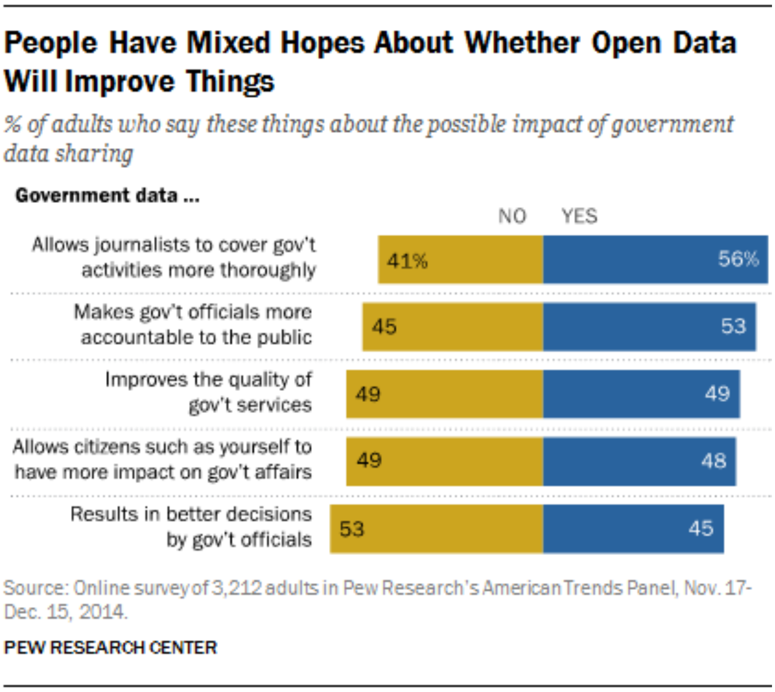



**Figure 3 - PEW Survey: Impact of OGD sharing.**

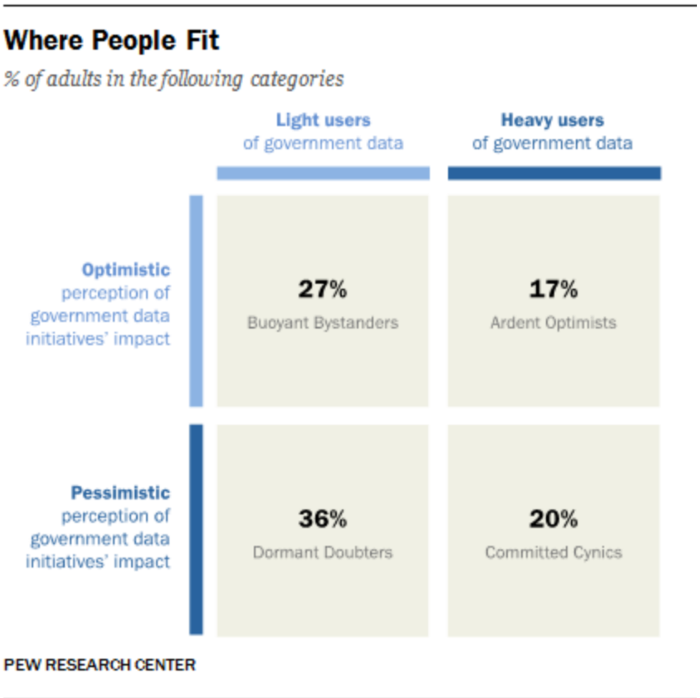



**Figure 4 - PEW Survey: Grouping Users of Open Data and Open Government Applications.**

Additionally, Figure 5 reveals that less than a third of citizens increasingly interested in acquiring government data. However, the percentage of people interested in getting info/data is higher than that concerned to routine operations as paying a fine or renewing a hunting and fishing license.

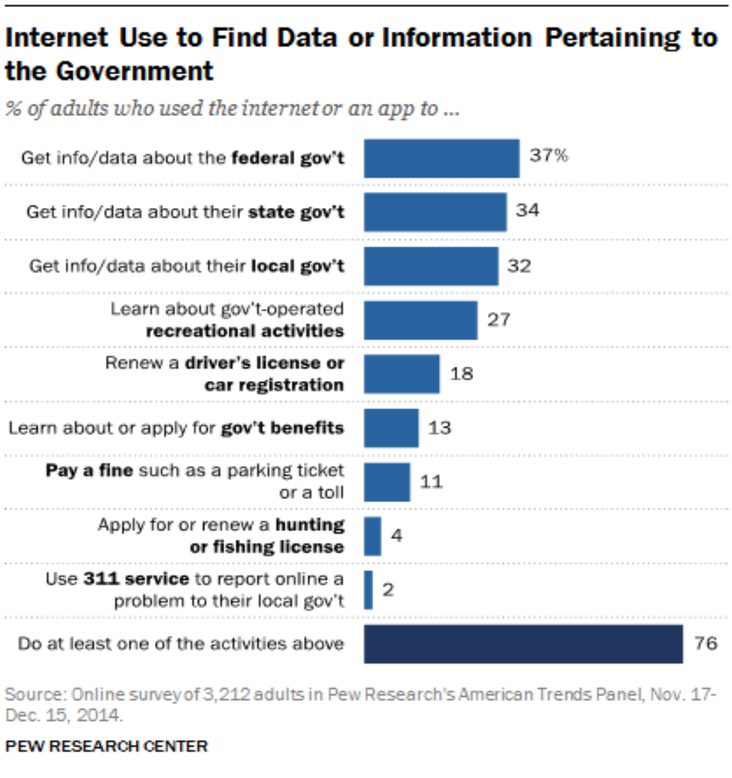



**Figure 5 - PEW Survey: Internet use to find data or information pertaining to the government.**

A complete analysis of these aspects is out of the scope of this thesis that tries to offer an understanding of the technical factors, which influence the emerging phenomenon of OGD. In particular, the next section focuses on technical aspects entailed by the eight principles of Open Government Data that prevent the adoption of an open data approach in government contests and public administration as devised by eight principles of Open Government Data.

## 1.2 Technical Factors Influencing the Opening of Government Data

Currently, several administrations store their data in platforms that capture the legacy data from public sector entities, support gathering and storing data from the government activities and organize datasets in order to enable easy downloading and sharing by users. Unfortunately, the principles in Table 3 are not legally required by many open government initiatives. Many governments only require publishing data and are slow to embrace the open data paradigm that potentially disrupts traditional government practices [34]. Even in government organization that aim to adopt advanced technologies and assess that the open data paradigm is relevant and necessary for improving citizens' engagement [35], many barriers of technical nature obstacle the "openness" of government data. In the follows, we address the question of what technical aspects, alone or often in combination, are most relevant in preventing the full application of open data paradigm.

1) STANDARDS FORMATS FOR PUBLISHING DATA

The fifth and the seventh principle assert that data must be non-proprietary and published in a machine-processing format. However, many of the efforts surrounding OGD concentrates on placing all allowable and available data on a web portal in a large variety of data formats that are often proprietary (e.g. such Microsoft Office formats, PDF, etc.) and hardly convertible to open standards such as XML and RDF. It is not a rare event that data encoding (often left on an agreed upon government agencies) affects the semantic of information. A common example is addresses that come in many different formats. Consider the following example:

- 678 3th Avenue Northeast
- 678 3 Ave NE

- 678 Third Avenue NE

Beside differences in their formats, the above addresses are the same location. A similar example can be given for temporal data (i.e. dates and calendars) largely used by financial agencies in the description of their data. Moreover, many government offices transfer datasets of addresses to each other, which are reformatted by specific software into the receiving office's specific format. Diversities in coding the same information make difficult to automatically capture temporal trends or geographic information by software applications. As a consequence, many data, although available in public portals, are semantically ambiguous and only partially useful [36][37].

2) INTEROPERABILITY

A recent white paper on interoperability [38] demonstrates that interoperability happens at different technical levels and describes two aspects of data interoperability:

*Syntactical Interoperability* - is usually associated with data formats. Certainly, the messages transferred by communication protocols need to have a well-defined syntax and encoding, even if it is only in the form of bit-tables. However, many protocols carry data or content, and this can be represented using high-level syntaxes such as HTML or XML. This aspect usually refers to machine-processing data that are not in a proprietary format.

*Semantic Interoperability* - is usually associated with the meaning of content and concerns the human rather than the machine interpretation of the content. Thus, interoperability on this level means that there is a common understanding between people of the meaning of the content (information) being exchanged.

Given the heterogeneity of datasets published onto portals, a popular approach for supporting data consumers is to collect metadata i.e. keywords used for describing online datasets and organizing data into a catalogue. Specifically, there are two types of metadata:

- Descriptive metadata (o external metadata) typically used for describing datasets in terms of their author, data of creation, keywords about their content, etc. Such information is used for discovery and identification purposes;

- Structural metadata (o internal metadata) are a mean for expressing a catalogue that organizes datasets into a logical structure in order to permit the data to be applied, interpreted, analysed, restructured, and linked to other, similar, datasets.

Essential for data searching, metadata allow interoperability between different systems by means of agreed-upon catalogues that can be automatically queried by external and unrelated software systems in order to capture and use remote data. Known as "harvesting", this capture originates interoperability issues when an aggregation of metadata is required [39]. Indeed, only if the metadata structure and catalogue are standard and/or self-explicative enough, it is possible to group together metadata from several data resources into a single coherent catalogue that integrates all various data offers.

But two central questions arise. The first question is: what is recorded in the metadata of OGD datasets and how? Usually, the quality of OGD's metadata is low (just dataset's name, description and author) and there exists a large heterogeneity in terms of semantics and standards for metadata catalogues. Beside the few proposes in literature [40], the lack of agreed-upon standards for metadata publishing in machine readable form with good quality metadata still results in many challenges to be overcome for harvesting OGD metadata in a way that is useful for data consumers.

The second question is: which is the level of "openness" of OGD portals? The Global Open Data Index [41] provides the most comprehensive snapshot available of the global state of open data. In particular, the Global Open Data Index collects and presents information on the current state of open data release around the world and is run by Open Knowledge International with the assistance of volunteers from the Open Knowledge Network around the world. The Global Open Data Index is not an official government representation of the open data offering in each country, but an independent assessment from a citizen's perspective. It is a civil society audit of open data, and enables government progress on open data by giving them a measurement tool and a baseline for discussion and analysis of the open data ecosystem in their country and internationally from a key user's perspective. The Global Open Data Index is not only a benchmarking tool, it also plays a powerful role in sustaining momentum for open data around the world and in conveying civil society networks to use and collaborate around this data. For example, if the government of a country does publish an open dataset, but this dataset cannot be found through a simple search and/o its content is not clear to the public and, then the data will be easily overlooked and not put to good use.

Governments and open data practitioners can check the Index results to see how accessible the open data they publish actually appears to their citizens, see where improvements are necessary to make open data truly open and useful, and track their progress year to year. According to the common open data assessment framework there are four different ways to evaluate data openness – context, data, use and impact. The Global Open Data Index is intentionally narrowly focused on the data aspect, hence, limiting its inquiry only to the datasets publication by national governments. It does not look at the broader societal context – for example the legal or policy framework, (FOI, etc.) – and it also does not seek to assess use or impact in a systematic way. Lastly, it does not assess the quality of the data. This narrow focus of data publication enables The Global Open Data Index to provide a standardized, robust, comparable assessment of the state of the publication of key data by governments around the world.

Figure 6 shows the resulting Global Open Data Index from [41].

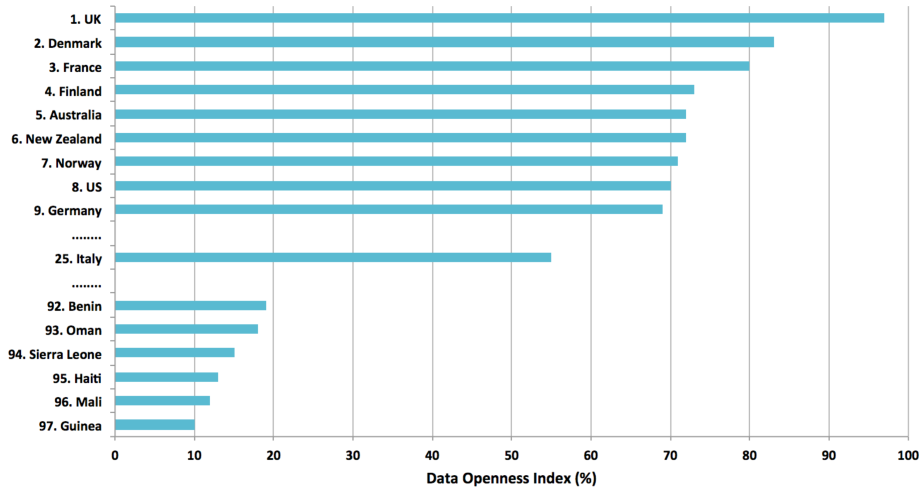
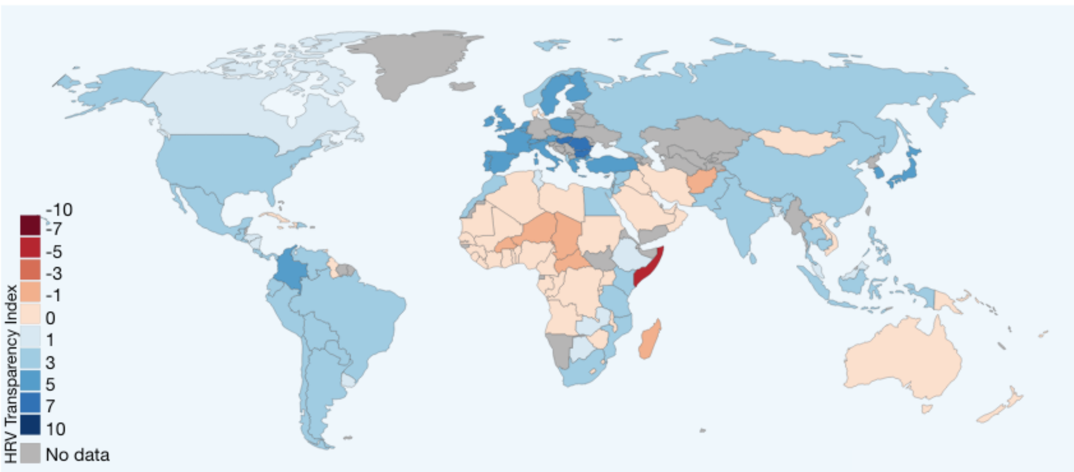



**Figure 6 - Global Open Data Index by Place, 2014.**

Beyond the openness, other closely related aspects to the open data have to be taken into account: the transparency and the impact in improving services for citizens and the effectiveness and accountability of public action.

Transparency is seen as crucial to government accountability, but it includes different dimensions with distinct effects and its measurement remains elusive. How to quantify the transparency, the first step to actually measure the expected improvements? In [42] has been defined the HRV transparency index (Hollyer-Rosendorff-Vreeland). This index is the result of a series of theoretical and empirical studies based on economic data for 125 countries from 1980 to 2010. The HRV index measures a specific aspect of government transparency: reporting national data to World Bank by means of international agencies, the project is still in progress [43]. Since the World Bank omits data considered "questionable", the HRV Index relies on a credible aggregate of economic data to form a measure of government transparency reproducible by others. This measure is a precise and narrow conception of transparency: the disclosure of policy-relevant information by the government to the public. Figure 7 shows a worldwide map based on HRV Index in 2010. As acknowledged by the authors, this map offers a narrow level measure of transparency.

Kosack and Fung [44] have distinguished four fundamental types of transparency: 1) The right to government information, embodied by laws such as the U.S. Freedom of Information Act (FOIA); 2) transparency around private organizations and corporate behaviour, often spurred by consumer campaigns; 3) government regulatory transparency such as "financial disclosures of corporations, product safety disclosures"; and, 4) "transparency for accountability," which can help improve delivery of services in fields such as education and health. From [44], Figure 8 correlates these four types of transparency with targets and users.




*(Data source: Hollyer, Rosendorff, and Vreeland (2014) – Measuring Transparency)*

**Figure 7 - HRV Government Transparency Index, 2010.**

Specifically, the authors focus their analysis on "transparency for accountability" because it reflects "the evolution of transparency from an end in itself, or an ingredient with important but nonspecific benefits for democratic governance, into a tool for dealing with increasingly practical and specific concerns of government performance." This area has increasingly become of interest for policymakers, foundations and other interested parties. In their work Kosack and Fung aim to provide a conceptual framework for political scientists, other scholars, and policy makers to understand the contemporary debates, as well as the emerging body of evidence, at the heart of the relationship between transparency and improved governance.

| | Users of transparency | |
|---|---|---|
| **Targets of transparency** | **Self-governing citizens** | **Individual customers/beneficiaries** |
| **Governments** | **I. Freedom of information** (e.g., use by journalists and citizens) | **IV. Transparency for accountability (T/A)** (e.g., disclosure to improve public services in health and education) |
| **Private firms/ corporations** | **II. Transparency for responsible corporate behavior** | **III. Regulatory transparency** (e.g., financial disclosures of corporations, product safety disclosures) |

**Figure 8 - Users and targets of transparency.**

Measuring the concrete impact of open data initiatives is a difficult challenge given its complexity. Since 2013, the measure the global and regional impact of open data on business, politics and civil society is one of objectives of the Open Data Barometer (ODB) [45]. It is a collaborative and global effort, led by the World Wide Web Foundation [46], which proposes a methodology that combines contextual data, technical assessments and additional indicators to estimate the evolution of open data impact [47]. Annually ODB identifies case studies in academic literature or in the media and uses them as a proxy approach to impact measurement. The ODB methodology is based upon three kinds of data: 1) a peer reviewed expert survey; 2) A government self-assessment in the form of a simplified survey; and, 3) Secondary data from Word Economic Forum, World Bank, UN e-Government Survey and Freedom House.

3) STANDARD AND USER-FRIENDLY PROCEDURES FOR QUERYING GOVERNMENT PORTALS

The easy of discovering and capturing OGD depends on several factors such as the accuracy of metadata and the access mode allowed by the OGD portal. Some portals only support simple search functions guided by keywords and/o just the download of their files. So, the results are often a plethora of heterogeneous datasets (e.g. tables, maps, images, documents, etc.) most of which are not of interest for the user who is forced to make a long and daunting selection [48]. This situation becomes worse if data are spread over a number of distributed web resources [49].

Additional difficulties come from data overlapping. Often, the same data is be published in several sites according to a vertical flow from local to regional and national organizations. For example, budget datasets from a state university can be published on the university portal, but also made available in a regional and in a national portal. This overlapping favours the duplication and the incoherence of data. According to the modern tendency of generating a big amount of data, the focus of public administrations seems to focus on making new information known, rather then on improving access data that was already published.

4) LACK OF USER INTERACTION

The widespread acceptance of WEB 2.0 and its results, such as social media, has created increasing demands and expectations about the use of such technologies in government organizations, the involvement of citizens in government actions and the delivering of more effective information in a way that respects security and privacy. Expectations are also about the development of applications with a high level of user interaction for capturing user attention. For example, a farmer may be not interested in downloading a large table about the loans granted to European farmers, but he could look with great interest at a Google-map showing the geographic distribution of these loans. Additionally, it should be even more interesting for him to see the distribution of the same loans in an area of the map that he selects interactively. Finally, he should be interested in sharing results with his friends via a social network.

The above example demonstrates that spatial information and social media are powerful means to improve the transparency and the openness of government data made available to citizen. As well, Web 2.0 technologies allow data collected by government to be "mashed-up" with data and services from several sources (including non-governmental organizations and private companies) for producing new insight about political decisions and new services for the private sector [50].

The great potential of Web 2.0 technologies is indisputable, but the related application should be evaluated responsibly by government organizations to protect privacy and security of the public. For example, data aggregations could originate a loss of transparency and accuracy or misleading citizens by providing a distort reality of facts. A recent research work [51] reviews the literature surrounding the use of Web 2.0

technologies in the context of e-Government and proposes a framework for the classification of the risk factors deriving from the use of such technologies. As showed in in Figure 9 and Table 5 these factors range from the technical aspects to the political and social aspects. The major barriers seems to be the diversity of transparency standards and cultural obstacles of different public institutions. As well, the interaction with political and institutional components [52] such as city council, city government, etc., and external pressures (political agendas and politics) might affect the process and results of IT initiatives. Recently, the application of Semantic Web technologies for the publication of open data government has an extremely high potential and impact [53] and can provide a tool for the unification and facilitation of data integration from multiple heterogeneous sources.

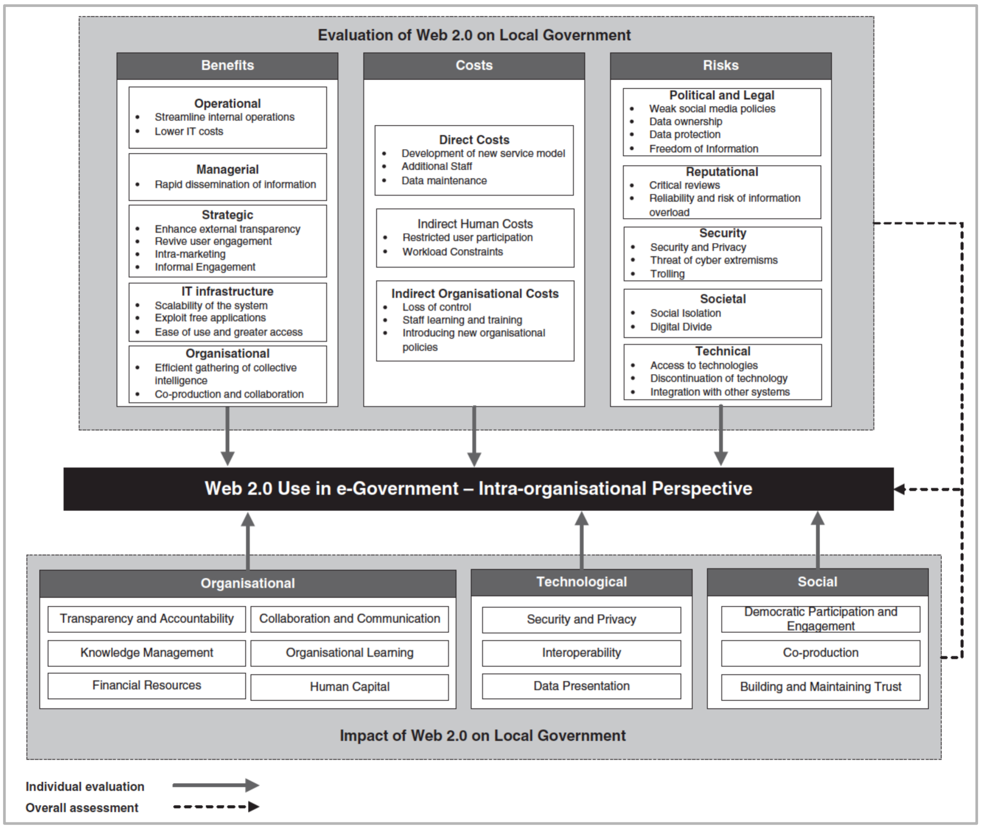



Figure 9 - An emergent framework for evaluating the use and impact of Web 2.0 in e-Government.

| Classification | Factors | Description |
|---|---|---|
| Political and legal | ▪ Weak social media policies | ▪ As Web 2.0 is an emerging phenomenon in government organisations some of the organisational policies governing the use of social media applications may still be in their infancy. Immature policies might prove to be a risk for government organisations. |
| | ▪ Data ownership | ▪ The technique of application mashups and content syndication on to existing e-Government platforms can also be an issue leading to loss of ownership control and authenticity of the final products. |
| | ▪ Data protection | ▪ Rise in responsibility for government organisations to handle personal information about individuals sensitively as most Web 2.0 technologies require this information to use the applications. |
| | ▪ Freedom of information | ▪ The use of Web 2.0 technologies can present challenges in appropriately responding to Freedom of Information legalities. It can raise significant issues for an organisation with regards to open access and the publishing of information. |
| Reputational | ▪ Critical reviews | ▪ While the advent of Web 2.0 technologies has played an important role in providing people with useful assessments of products and services, it has also meant that there is now a greater risk of these assessments damaging the image of people and organisations without good reason. This is because it is difficult to ascertain if assessments are fair or the result of personal resentment |
| | ▪ Risk of information overload and reliability | ▪ There is a risk of information overload and poor quality content shared with public users when using some Web 2.0 applications such as blogs and wikis, as concerns can be raised about their reliability, accuracy and authority |
| Security | ▪ Security and privacy | ▪ The open nature of Web 2.0 presents significant challenges to the traditional enterprise approach to controlling intellectual property over information shared and security of these applications. |
| | ▪ Threat of cyber extremisms | ▪ These new, interactive, multimedia-rich forms of communication provide effective means for extremists to promote their ideas, share resources, and communicate with each other |
| Societal risks | ▪ Social isolation | ▪ Though Web 2.0 can stimulate social interactions and communication between different individuals, there is also the risk of people isolating themselves from the real world as they become too addicted to the use of the internet |
| | ▪ Digital divide | ▪ There could be a risk of inequality between different groups of users in terms of access to, use of or knowledge of Web 2.0 applications. Some users may be hesitant of using Web 2.0 technologies and may not be interested in using the applications at all. This could indirectly result in the exclusion of these users and not allowing for equity of access. |
| Technical | ▪ Access to the technologies | ▪ The need for minimum requirements such as a device and internet access at a speed sufficient to support social media content |
| | ▪ Discontinuation of technology | ▪ The risk of the continuity of existing Web 2.0 applications. For example Yahoo announced the discontinuation of its Delicious tagging service. |

**Table 5 - Classification of the risks of Web 2.0 in the context of e-Government.**

The LOD initiative has been widely adopted and is now considered the reference practice for sharing and publishing structured data on the Web [54]. Since cities usually have large amount of heterogeneous data, Semantic Web best practices are key drivers in the definition of data reengineering, linking, formalization and use. Authors in [55] presented a methodology used to collect, enrich, and publish LOD for the municipality of Catania, a city on the east cost of Sicily. The collected city data, their production process, issues faced in creating a semantic data model and tools used for ensuring semantic interoperability during the transformation process were presented along with discussions/suggestions on how to use the produced open data model by local stakeholders (developers and main actors). Work in [55] has been carried out within PRISMA, an Italian project whose goal was to develop an "Italian cloud" for the needs of PAs that might be sustainable over time and to let LOD be exploited by small and medium enterprises in order to create new applications capable of delivering services tailored to the specific needs of citizens.

## 1.3 Actors and Roles in OGD Management

As previous mentioned, OGD is a philosophy – and increasingly a set of policies – that promotes transparency, accountability and value creation by making government data available to all. Public bodies produce and commission huge quantities of data and information. By making these datasets available, public institutions become more transparent and accountable to citizens. By encouraging the use, reuse and free distribution of datasets, governments aim to promote business creation and innovative, citizen-centric services.

This section explores how the disclosure of government data (GD) affects the interest of the different actors and ultimately disrupts the traditional roles of producers and consumers of GD. The identification of these actors and their roles is important both for better understanding how the open data paradigm impacts the government agencies and for highlighting the potential challenges in sharing and integrating OGD.

Government data originates from the public employees that encode and store information required by government processes into government information systems. This activity creates the primary sources of OGD and consists of elementary tasks such as storing administrative data about the budget or more complex actions, such as organizing the updates on a piece of pending legislation. Data definitions and encoding techniques change within the government departments and greatly influence the usability of data, its comparison, integration and fitness for use. These issues, although important, were not so relevant before the advent of OGD because the intended users of the primary sources (e.g. data files, databases, platforms, etc.) were public employees and persons interested in the processes of civil government: data and government information system have been created for them and applications, websites were often customized in order to meet their objectives and needs. Originally, the actors involved in producing and managing government data were public employees and stakeholders that play both the roles of data producers and data consumers. Nor do these actors have a good understanding about why and how to integrate several primary sources of GD, combine government data with non-government data owned by external providers, share information with citizens and stakeholders outside the traditional boundaries of the public administrations. Advances in ICT (such as Web 2.0. and Open Data) and the advent of open government data paradigm have transformed the above traditional scenario in a new context: here opening data involves not only the internal management of government agencies but, most important, external stakeholders, such as citizen and private organizations, that are consumers and/or producers of data.

Indeed, new technologies make possible the integration of the primary sources of GD (previously owned by government agencies) with external data such as external files, geo-coded data, social network data and data crowded by citizens. This integration results in applications that change the way in which original government information is organized and brings out a scope of the data, which differs from that originally planned by the data owners.

Within their role of data providers, both government agencies and external stakeholders are now in charge of enabling the acquisition and assessing the accuracy of data they own. In turn, data is combined and managed as a resource for creating applications that impact with legal, organizational and social factors because their intended users are often persons or groups whose interests change over time and often are not aligned with the interests of the data owners. This vision of data as a resource encourages good data management practices in the government agencies, efforts in making data available in machine-readable formats, effectiveness in trying to make data easy to re-use.

The production of fit-for-reuse data requires efforts that are higher than those required to place together a variety of information into web sites that appear as monolithic silos of heterogeneous and often unstructured data with increased costs and efforts to open data afterwards.

Although government agencies vary their efforts and levels of effectiveness in providing open information, it is not enough to focus only on the human factors but open data initiative should also consider the adoption of advanced technical solutions as an important factor in trying to accomplish OGD disclosure with minimum effort and cost.

# 2 Publishing and Consuming OGD

Particular efforts have been undertaken with a large number of countries to increase the accessibility of OGD and a variety of technical approaches exist.

Inspired by the original paradigm of the Web, the first generation of Open Government Data Portal (OGDP) had the purpose of simply making OGD available to users [56]. The advent of the Semantic Web and the Web 2.0 contributed to gathering and storing data from the public administration activities and other domains while facilitating their accessibility and reuse. Today, OGDPs are considered not merely the medium through which open government datasets are made accessible to the public but a vital technical support for assembling the legacy data from various administrations and organizing them in a manner that supports easy downloading, modification and sharing [57]. ICT advances, new governance arrangements and new user practices result in new ways to bootstrap the value generated by OGD.

This section underlines the basic aspects of this bootstrapping process by presenting the advanced technical solutions for publishing data and some value-added applications that relies on OGD for creating services of interest for the citizens.

## 2.1 Technical Solutions for Bootstrapping OGD

Currently, the basis of OGD initiatives is gathering large amount of data with their respective metadata and publishing them in an Open Government Data Portal. The most adopted approach is the publication of metadata catalogues on stored data although difficulties arise in harvesting metadata records and facing the lack of semantic interoperability due to the incoherence between metadata structure and meaning.

Towards mitigating this problem, advanced software platforms (usually denoted as Open Data Platforms (ODPs) and catalogues/vocabularies are technical solutions suited for increasing the “openness” of data. Often, both these two solutions are adopted. In this case, the ODP manages OGD datasets and interfaces the catalogue that, in turn, stores and manages the metadata of all datasets, documents and applications. Searching is enabled by displaying the content of the catalogue.

The term Open Data Platform does not have a universal definition because it is a relatively new concept still under development and not much research and conceptualization have been done on this field. However, the term “Platform” represents a system defined by three aspects: (1) a stable, low-variety “core”, (2) a changeable, high-variety set of “complements”, and (3) the interfaces which allow core and complements to operate as a single system [58]. The platform architecture is a related concept defined as “a conceptual blueprint that describes how the ecosystem is partitioned into a relatively stable platform and a complementary set of modules that are encouraged to vary, and the design rules binding on both” [59].

Despite the rapid research and development in this area, the ODP technology is still in its infancy. Most of the existing open data platforms can be viewed as cataloguing system for open data; they have been extremely useful in kick starting easy publishing of large volumes of open data in diverse data types. But the raw nature of data being shared on these platforms makes it hard for ordinary users to effectively exploit the data shared on these platforms. As well, advanced skills are required to transform the data in a manner in which they can easily exploited for analysis and discovery purposes. Existing open data solutions are missing proper easy to use workflows for extracting and transforming data in machine-readable formats and offer searching, querying, harvesting, visualizations of data with limited support for the analysis of data. Some platforms have exploited sematic web technology and advance indexing techniques to enable easy integration and exploitation of open data across datasets and portals. Data discovery, fine grain searching, advance analytics and Q&A over open data still remain challenging features to make open data platforms useable for ordinary users. API and external tools are normally used to developed applications.

The support of geospatial data standards and tabular formats (such as CSV, excel, etc.) is much better than other formats in most available open data platforms. Basic visualization and analytics being offered by open data platforms is satisfactory. Support for customization, personalization, access control and other configuration features vary across different platforms. DCAT [60] is supported by majority of platforms as format for metadata exchange. Collaboration and sharing is supported widely, either as internal solution or as an extension to platform that have community editions with technical support for their extensions. The tools and technologies used for the development of open data platforms are quite ubiquitous and easy to learn. In general, the documentation provided by most of the platforms is well formed and satisfactory.

Although focusing on National Statistic Offices (NSO), a recent report from World Bank Group [61] devises the following assessment criteria for evaluating the quality of the ODP and the management of OGD:

**Descriptive metadata**: Corresponds to external metadata and are typically used for discovery and identification. Related information is about title, author, subjects, keywords, publisher.

**Machine-readable**: Data available as machine-readable structured data in a non-proprietary format can be easily read and used by software systems without human interpretation.

**Anonymous access**: Users can search for and access data and metadata without having to identify themselves, create a user account, or receive advance permission.

**Data licenses**: Data licenses (terms of use) associated with each dataset are clearly presented to the user and permit reuse and republication of that data in any alternative form.

**Data attribution**: Users can cite, attribute, and link to datasets, and contact data owners if they have questions.

**Search**: Search results should return focused summaries on datasets, along with keywords, which aid classification, and the option of reviewing the data online to assess its content.

**Application Programming Interface (API)**: Platforms make their contents available to external systems by supporting programmatic queries and access to metadata and resources.

**Uniform Resource Identifiers (URI)**: Platforms make datasets available at persistent URIs that never change, allowing them to be externally referenced reliably.

**Harvesting**: An automated and autonomous mechanism for capturing data from known web addressable locations into a single database or data repository.

**Federation**: A meta-DBMS, transparently maps multiple autonomous database systems into a single federated database, allowing discrete data publishing systems to be integrated, yet operate independently.

**Public Documentation**: Data platforms provide comprehensive information for developers and the general public on how their platform works. Such documentation should be updated with each new software release.

**Standards-based**: Platforms are consistent with emerging standards recognised by the W3C especially as regards metadata, RDF, and hyper-cubes.

**Data Endpoints**: Structured data endpoints return data in predictable ways ranging from a known type of serialisation format to more complex implementations allowing the data to be queried, filtered or refined prior to their download.

**Visualisation**: Availability of tools to present data as common charts, maps or perform more complex statistical analysis.

**UX & S/W Extensibility**: Sufficient template and layout customisation to provide a consistent user-experience and provide a common look and feel across all online services.

According to the above criteria, the report also presents an evaluation of the most commonly used software platforms relative to the above assessment criteria (Table 6) where all platforms seems to share a dark side: they lack a good support for harvesting and federation e.g. for enabling system interoperability. Five software platforms seems better meet the stated criteria i.e. CKAN, DKAN, Socrata, Semantic Media Wiki and OpenDataSoft. Existing open data platforms have been compared by ROUTE-TO-PA, a recent Horizon 2020 project [62] that combines expertise and research in the fields of e-government, computer science, learning science and economy. The project aims at improving the impact, towards citizens and within society, of ICT-based technology platforms for transparency.

| Software platform | Open Data-Specific Criteria | | | | | | | | | | | | NSO-Specific Criteria | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Descriptive Metadata | Machine-readable | Anonymous access | Data licences | Data attribution | Search | Open API | Static URI | Harvesting | Federating | Public Documentation | Standards-based | Structural Metadata | OLAP Hypercubes | Data Endpoints | Visualisation | UX & S/W Extensibility |
| CKAN | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ⬩ | ⬩ | ◆ | ◆ | | | ⬩ | ⬩ | ◆ |
| DevInfo | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | | | ◆ | | ◆ | ◆ | ⬩ | ◆ | ⬩ |
| DKAN | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ⬩ | ⬩ | ◆ | ◆ | | | ⬩ | ⬩ | ◆ |
| Junar | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ⬩ | ⬩ | ◆ | ◆ | | | ⬩ | ◆ | ◆ |
| NADA | ◆ | | ⬩ | ◆ | ◆ | ◆ | | ◆ | | | ⬩ | ◆ | ◆ | | | | ◆ |
| Nesstar | ◆ | ◆ | ◆ | ⬩ | ⬩ | ◆ | | ◆ | | | ⬩ | | | ◆ | | ◆ | ⬩ |
| OpenDataSoft | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ⬩ | | ◆ | ◆ | | | ◆ | ◆ | ◆ |
| PC-Axis & PX-Web | ◆ | ◆ | ◆ | ⬩ | ⬩ | ⬩ | ◆ | ◆ | ⬩ | ⬩ | ⬩ | ⬩ | ◆ | ◆ | ◆ | ◆ | ◆ |
| Prognoz | ◆ | ◆ | ◆ | ⬩ | ⬩ | ◆ | | ◆ | | | | | | ◆ | | ◆ | ⬩ |
| Semantic MediaWiki | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | | | ◆ | ◆ | | | ⬩ | | ◆ |
| Socrata | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ◆ | ⬩ | ⬩ | ◆ | ◆ | | | ◆ | ◆ | ◆ |
| Swirrl | ◆ | ◆ | ◆ | ◆ | ◆ | | ◆ | ◆ | ⬩ | ⬩ | ◆ | ◆ | | | ⬩ | | ◆ |

◆ offers a complete solution ⬩ offers a partial, or incomplete, solution.

**Table 6 - Summary and evaluation of Platform Features**

The deliverable D2.1 of ROUTE-TO-PA [63] presents a report about the state-of-the-art and evaluation of existing Open Data Platforms. From this report, Table 7 shows a comparison between the five platforms that better meet the assessment criteria in [61]. Moreover the report presents also an evaluation about the availability of extensibility mechanisms in ODPs (Table 8).

| Features | CKAN | DKAN | SOCRATA | ODS | SMWIKI |
|---|---|---|---|---|---|
| Data, Metadata & File Format Standards | ● | • | ● | ● | • |
| Search & Indexing | ● | • | • | ● | ● |
| Social Media, Sharing & Collaboration | ● | ● | ● | ● | ● |
| Publishing Workflow | ● | ● | ● | • | • |
| Harvesting, Federation & Catalogue | ● | ● | ● | ● | • |
| Data Analysis | • | ● | ● | • | × |
| Visualization | • | • | ● | ● | • |
| Personalisation | ● | ● | ● | ● | ● |
| Customisation | ● | ● | ● | • | ● |
| Licensing for Dataset | ● | ● | ● | ● | ● |
| Accessibility | ● | ● | ● | • | ● |
| Extensibility | ● | ● | ● | • | ● |
| Technical Environment | Python | PHP, Drupal | Scala | N/A | PHP |

● / • / × extensive / limited / not provided solution

Table 7 - Comparison of five selected platforms.

| Platforms | Extensible | Open Source | Extension Mechanisms | Guide Available | Customisable Metadata |
|---|---|---|---|---|---|
| CKAN | ● | ● | ● | ● | ● |
| DKAN | ● | ● | ● | ● | ● |
| Socrata | • | × | • | ● | ● |
| Open Data Soft | • | × | • | ● | × |
| Semantic MediaWiki | ● | ● | ● | ● | ● |

● / • / × extensive / limited / not provided solution

Table 8 - Availability of extensibility mechanisms in ODPs.

Although the new assessment criteria adopted by ROUTE-TO-PA, the two evaluation seems to agree apart from 1) the evaluation of OpenDataSoft that is much worse; 2) limitations in the availability and the extensibility mechanisms of Socrata; 3) lack of support for data analysis in Semantic Media Wiki.

To recapitulate, CKAN and DKAN standout from the software platforms for open data as their features satisfy almost all the requested assessment criteria.

Being interested in OGDP, this thesis focuses on CKAN for the following motivations. CKAN is a complete open source platform widely used in Europe [65] and recently in US [18] by about 150 governments organizations and communities around the world [68]. Because of this broad use and its high alignment with DCAT [64] (the most prominent data catalogue vocabulary), CKAN is the de-facto standard for metadata catalogues [40]. The following sections outline the basic features of CKAN and DKAN offered by the respective official sites [66][69].

### 2.1.1 CKAN

CKAN (the Comprehensive Knowledge Archive Network) is an open source data portal platform developed by Open Knowledge Foundation (OKFN), a not-for-profit organization created in 2004 "to promote the openness of all forms of knowledge" [66]. One of the strengths of the open source model is in the communities that form around free software products. The CKAN community is no different, and is arguably one of the strongest open data communities in the world. Together, the CKAN community has a wealth of knowledge and expertise that other people using the CKAN software can draw on. The Open Knowledge Foundation draws on and contributes to this rich resource to help developers to drive CKAN product development.

CKAN is built with Python on the backend and JavaScript on the frontend, and uses the Pylons web framework and SQLAlchemy as its ORM. Its database engine is PostgreSQL and its search is powered by SOLR. It has a modular architecture that allows extensions to be developed to provide additional features such as harvesting or data upload. CKAN uses its internal model to store metadata about the different records, and presents it on a web interface that allows users to browse and search this metadata. It also offers a powerful API that allows third-party applications and services to be built around it.

Overseen and managed by the CKAN Association [67], the platform is specially suited for government agencies, organizations and companies who want to publish and share open data [68]. The original aim of the CKAN's developers was to create "an open-source DMS (Data Management System) for powering data hubs and data portals". From its first public release in July 2007, CKAN has evolved over time from a simple registry and catalogue of distributed datasets to a data repository for the upload of data 'blobs' or files (version 1.6 released in 2011) to a datastore (version 1.7 released in 2012) for the management and the visualisation of structured data. The current version of CKAN (2.5, March 2016) is extremely versatile: it is possible to use separately the catalogue, the repository and the datastore or, in alternative, combine the use of all these components.

CKAN provides a streamlined way to make data discoverable and presentable as it allows the publication of data via import or through a web interface. It enables storing the raw data, metadata and structured data that can be visualized with interactive tables, graphs and maps. As well it allows for searching geospatial data on a map by area. The API's rich programming interfaces benefit from over 60 extensions and include link checking, comments, and analytics. Most important, users are enabled to build communities of federate networks with other CKAN nodes and they can comment on and follow datasets.

### 2.1.2 DKAN

DKAN is an open data platform that is based on Drupal and maintained by NuCivic [69]. It provides a full suite of cataloguing, publishing and visualization features that allow governments, non-profit organizations and universities to easily publish data to the public. Built with supports and inputs from OKF, DKAN seeks to replicate CKAN 2.0 functionality, standards and API configuration; and does, in fact, reuses CKAN components wherever possible [71] [61]. There is however, a point of difference between CKAN and DKAN. Indeed, being a distribution (pre-configuration) of Drupal, DKAN is a complete CMS offering comprehensive tools to manage content, documents, and community, in addition to datasets which is presumably impossible in CKAN [61].

A summary of the features of DKAN is offered by [61] and it is presented below. DKAN imports and interprets datasets in CSV, XLS, XLSX and PDF file formats and also text files in a machine-readable format. As a current shortcoming, DKAN renders data to users in the same format as it obtains datasets from publisher without any data transformation. DKAN has a clear and thoroughly documented online (but complex) API to download resources with output available as JSON or XML. DKAN harvests existing data resources and is able to regularly update streaming data, via the API. However, there is currently no user-interface for setting up automated harvesting tasks. Federating is made possible through DKAN's interconnections with Drupal. As part of standardization policy, DKAN is aligned with best practice in the open data industry, yet offers no support for metadata for data structure. DKAN's visualisation tool is described as 'public facing visualisation library', limited in support because it does not permit functionalities to *save or share* of specific visualisation materials. A new set of tools developed recently supports embedding and saving charts, including geospatial data, as part of data-driven initiative. Integration toolkits were developed to facilitate integration with third-party data visualisation web services such as CartoDB [72].

## 2.2 The Value-Added Applications

The reuse of OGD targets the development of applications and services of interest for the citizens with the aim of promoting the economic development. By 2011, cities, rather than federal government, bootstrapped the use of OGD by promoting contests with cash prizes to encourage the implementation of civic apps. However, the initial enthusiasm behind the development of OGD-based apps had waned and few applications were created and widely used. The report [73] surveys more than 350 existing apps (by 2012) in 13 applications domains, denoted as categories (Figure 10). The majority of apps target categories such Entertainment and Transportation with only a limited interest in exploring the primary sources of government and civic data. The same report underlines that the vast majority of the developed apps relies on a single dataset and only a few of them integrates more than one datasets. Some apps integrate OGD with data from the social Web or from Wikipedia. This finding indicates that apps 1) use only a small range of the available OGD datasets with a large overuse of certain datasets 2) try to combining user opinions to rank some public services such as parks and theatres.

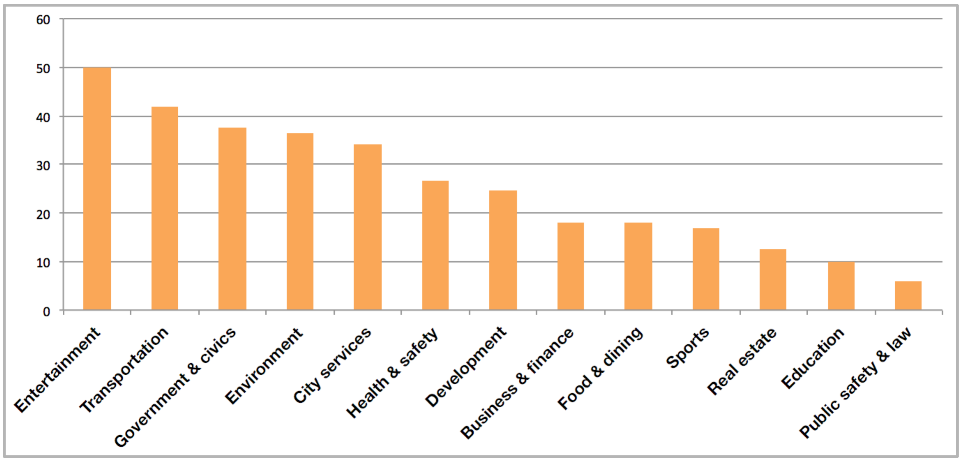



**Figure 10 - Number of OGD apps per category (2012).**

Apps are usually developed by freelancers and research institutes and are available for free. Rare apps are available in App Store and Android market and require a monthly subscription fee. This evidence indicates that the majority of open data initiatives do not interest the business community that seems to be not convinced about the valuable nature of OGD and its potential in fostering novel business applications.

An important feature of the examined apps is data visualization: all apps make great use of graphs, maps and tables for delivering data. In the same vein, government and civic apps visualize information in order to bridge the gap between citizen and the large amount of data stored in government sites.

Relying on GPS coordinated or postcodes, many apps are context aware (i.e. they identify the user's location) and use geographic maps from Google Maps and Open Street Map for location-based services. Few apps access OGD real-time as the most of apps use statistic datasets made available in several domains such as traffic, environment, transportation, crimes, etc.

Finally, apps are prevalently classified under civic services for engaging citizens in the reporting of some city issues such as street problems or excessive garbage. This final consideration is shared by a recent paper [74] as it assert that cities, rather than the federal government, took control of promoting open data initiatives. Moreover, the paper classifies the developed apps into two generations according to their characteristics. In particular, the first (and early) generation of apps fail in facing the open data challenges for several factors including the limited adoption and support by government, the publication of data with no commensurate changes in city services, the intensive use of certain datasets, the resistance to data transparency by public administration.

Relying on the lesson learned, the second generation of apps focuses on devising new mechanisms and new actors as well as on increasing the direct participation of city administrators. On-line marketplaces provide a venue for common apps and crowdsourced data repositories in order to engage communities and interested people such as developers and private citizens.

Developers with a limited experience with the civic services were embedded in city organizations for substantial time periods in order to increase both their knowledge about operative challenges and reciprocal engagement with the civic stakeholders. This interaction between the civic department and the developers contributed to bridge the gap between city services and external open data initiatives. Most important, the second generation of initiatives has encouraged the public administrations to adopt best practices for opening data repository to the public and the provision of data in usable formats. Such best practices include the efforts for promoting standardization of data formats and APIs to allow effective data sharing and entail coordination and procedural innovations that involve both technical and politic aspects.

DontEat [75] is a popular app that relies on the inspection data to the restaurants provided by the New York City Department of Sanitation. When patrons enter a restaurant, DontEat recognizes the local and lets them check its inspection status by sending an alerting message if infractions occur. Being made aware of violations, usually patrons leave the restaurant after the reception of the message. Because this behaviour has encouraged the restaurant owners to be compliant with the law, health inspector began to complete their work quickly and without multiple visits.

With similar purposes, the Department of Health and Mental Hygiene in USA released the app ABCEats [76], also available on iTunes that allows users to check inspection letter grades at restaurants near to their current location. The information is updated daily.

In 2009, the City of Edmonton, Canada, launched an Open Data Catalogue including information about street construction project. In the city website, this information was presented both in static and interactive maps. Specifically, static maps show projects located by numbered dots that reference a list of projects on the page below the map. When citizens click on a numbered dot, the interactive map pops up a detailed description of the project. Using this open data, a local application developer created a mobile app for accessing the map interface. He geo tagged original data in order to provide a higher level of detail. As a result of these efforts new data were included in the catalogue and new apps were developed. Currently, the city website allows to download several apps including ETS Live To Go, MyEPark Parking and SmartTravel [77].

ETS Live To Go provides customer-focused, safe, reliable and affordable services about real-time information for Edmonton Smart Bus routes. The user can use his current location to see real-time departures, save his favourite bus stop, address or landmark, access his personal contacts to find nearby bus stops.

MyEPark Parking App allows users to pay for their parking sessions, add money to their account and see their account balance without cards or coins. An account is required in order to use this application. Instead of downloading the app, users can rely on the one-time quick pay option on their desktop or mobile browser. Accessed by the city website this feature doesn't require an account and is ideal for starting or extending the parking stay, anywhere, anytime.

SmartTravel app gives drivers verbal warnings and the maps of school zones, speed limit changes, and high-collision locations. It also provides real-time warnings of: weather factors (freezing rain), traffic disruptions, seasonal considerations (school is back in), and enforcement (Big Ticket Event). The app meets all the requirements of the distracted driving legislation.

In the vein of experimenting citizens' interaction, Boston's Mayor's Office of New Urban Mechanics promoted the crowd-sourcing project Street Bump [78] that helps residents improve their neighbourhood streets. Volunteers use the Street Bump mobile app to collect road condition data while they drive. Boston aggregates the data across users to provide the city with real-time information to fix short-term problems and plan long-term investments. In partnership with New Urban Mechanics, Connected Bits designed and developed the app, collaborating with the design company IDEO. The City of Boston will make the app freely available so others can use and build on the project's efforts.

Focusing on citizen interaction, FixMyStreet Platform [79] allows to easily launch a website that helps people to report street problems like potholes, broken street lights, graffiti, fly tipping, broken paving slabs, etc.) without worrying about the correct authority to send it to. Problem reports are then sent to authorities for fixing and FixMyStreet takes care of that using the problem's location and category, and sends a report, by email or using a web service, to the department or body responsible for fixing it. But FixMyStreet doesn't just send problem reports : it makes the reports visible to everyone. Anyone can see what's already been reported, leave updates, or subscribe to alerts. Currently, there are 13 FixMyStreet sites all over the world in Uganda, Malaysia, India, Australia, France, Ireland, Norway, Spain, Sweden, Switzerland, United Kingdom, Chile and Uruguay.

Although not correlated with OGD, Roadify [80] is a notable example of civic app that aggregates and distributes real time transit arrivals, schedules and service alerts for display on any type of screen. Riders can easily find out what's going on, whatever they're doing wherever they are. Currently, Roadify counts 192 installations by cities over the world.

OpenSpending [81] is a notable example of specialized OGD data hub that hosts a free and open database of government finances worldwide to offer information on public spending, government money distribution, public financial transactions and tax values. This centralized open platform enables anyone to explore, visualize and track government spending. At time of writing this thesis, the OpenSpending database hosts more than 1000 datasets with over than 28 million of governmental financial transactions. The current Open spending platform originates from the joint efforts derived by the launch of the app “Where Does My Money Go? [82] (launched in 2009 for supporting citizens in exploring how governments spend the money collected from taxes) and similar initiatives carried on by several projects. Nowadays, Open Spending supports over 30 versions of “Where Does My Money Go?” worldwide.

Although numerous applications exist, the majority of apps have been produced by individuals and by civic administrations to provide cost savings and better services to citizens. In experiencing the use of apps, civic administrations have learned the importance of enlarging the sphere of distribution of their data and supporting the effort of crowd-sourcing content from citizens.  As well, the combination of OGD with maps, emphasizes the importance of geospatial data, allows a better understanding of the data in question while making data really useful. The most of apps uses a single data set, rather than mixed data. This is a serious setback that prevents to take full advantage from the real power of OGD that results more apparent when multiple data sets are used.

# 3 Cloud Technologies and OGD-based Applications: a Reference Architecture for Contextualizing Metadata

Even though OGD portals could be considered as a technology that supports open government initiatives, they are only at the initial step of the whole public accountability process and a number of drawbacks prevent them for completing this process. The previous sections have identified that most common challenges faced with the management of OGD and also the proposed solutions. Despite the apparent success and high impact of open data portals, Yu and Robinson [10] note that these initiative focus “more on technological innovation and service delivery” and public agencies “have tended to release data that helps them serve their existing goals without throwing open the doors for uncomfortable increases in public scrutiny.”

If publishing open data is one side of the coin, then making data discoverable is the other. Governments around the worlds have recognized benefits of disclosing internal data but lack initiatives for designing and implementing OGD-based applications [83]. However, the use of emerging technologies still requires significant additional effort towards the development of applications that serve citizens’ needs by harnessing OGD [84] although the Obama’s administration commitment to open government.

This chapter aims to investigate the extent to which Cloud technologies offer a viable platform for making OGD “fit for use” by various audiences including citizens and private stakeholders interested in developing and deploying value-added applications. The first section presents the architecture and the functionality of the cloud environments to give an idea about how cloud platforms could offer added-value service components and why the flexibility of cloud-based applications makes their adoption attractive for OGD.

The second section outlines the basic motivations for this thesis. The finale section introduces a framework that aims to be a reference architecture for the development of OGD apps.

## 3.1 Cloud Environment

The aim of a Cloud environment is to endow users with a service environment for the provision of compute power, storage and software. In particular, Cloud services refer both the software applications delivered as services over the Internet and the hardware and system components [85]. Cloud resources can be organized into the following bottom-up layers, depending on the type of service:

*Infrastructure-as-a-Service* (IaaS): provides suitable computer and storage hardware, including physical servers and network bandwidth services directly to end users which do not have to 'own' enough processing power, storage capacity, applications to meet their needs. These resources can be 'rented' from a cloud provider on an as-needed basis and accessed from anywhere via an Internet connection. As cloud services employ a metering system that divides the computing resources in appropriate blocks, users pay for the service as an operating expense without incurring any significant initial capital costs.

*Platform-as-a-Service* (PaaS): provides a hosting platform in which users are allowed to develop applications without dealing with software issues such as operating system upgrades, implementation of compilers, installation and maintenance of databases, etc.

*Software-as-a-Service*, (SaaS): it basically means that the software is available on demand, most of which is browser-based and devoted to a specific function. SaaS applications are often priced as a rental fee, which includes the application software license fees, software maintenance, and technical support costs.

Although provided at different level, Cloud Services communicate with each other and are self-contained i.e. each service provides the same functionality, independently of other services. The global infrastructure offered by a Cloud environment is made possible by abstracting the physical resources (i.e. storage, memory, network etc.) so that multiple operating systems can run on a single hardware platform concurrently. Commonly referred as virtualization, this abstraction grounds on technologies that greatly improve resource utilization and offer several key advantages, such as lower entry cost to use compute-intensive resources available only to the largest scientific organizations, pay-per-use utility model, high scalability dynamically adjusted according to user demand. The cloud computing platform composes of thousands or even tens of thousands nodes, which provide mass data storage, management and processing. Because these dynamically scalable and virtualized cloud resources are provided as services, users are not required to have expertise in, knowledge of or control over the technological structure that supports them.

As an innovative solution for structuring computer resources, Cloud Computing could offer many advantages to OGD-based applications that previous solutions do not have. In fact, the primary difference is that offerings and services are not located in house, but outside the open government portal. Accordingly, government organizations could avoid the installation and management of software on their own computers and further benefit from centralized and automatic software update in developing applications without significant hardware and software expenses or management time as costs directly reflect the amount of use.

The IaaS layer is well suited for the deployment of data intensive applications that benefit from functionality for getting as many machines as the analysis of data needs and automatically scaling up and down the hardware resources based on dynamic workloads. Conversely, in distributed processing environments, it is difficult to distribute and co-ordinate a large-scale job on different machines, run processes on them and install additional machines to recover if one machine fails. Cloud technologies remove such technological concerns from the users and promote applications such as MapReduce a run time system that automatically partitions input data and schedules the executions of programs for large datasets in a large cluster of commodity machines [86]. As well, its open-source implementation Hadoop has been seen widespread adoption in many application domains.

A less explored level of cloud services is PaaS which focuses on building applications such that draw the necessary resources on-demand (like compute servers or storage), perform its tasks, then relinquish the unneeded resources after a task is done. In a few words, a PaaS application scales up or down elastically based on actual need for resources. Everything is automated and operates without any human intervention by taking advantage of simple APIs of Internet-accessible services that scale on-demand. A relatively few number of applications exist which have been developed at PaaS layer, beside PaaS based solutions should be an added value for the developers as they allow any developer to deploy web applications using software libraries and platforms by the cloud provider. In particular PaaS applications are attractive for small organizations, such as single developers or small enterprises, which are allowed to benefit from the most advanced technology available today at lower costs and lower risk. Indeed, the upfront investment is low, they don't need to invest in software / hardware, licensing and renewal costs are kept to a minimum and they only pay for what they use.

Cloud based applications contrast sharply with traditional centralized computing platforms that are not only costly, but risk to become increasingly inadequate to meet the application requirements. As reported by [87], a significant number of research institutions are experiencing the capacity limits of its computing facilities. Additionally, collecting and configuring suitable tools and resources for certain research purposes is non-trivial job, even for expert developers. As such, PaaS based solutions can offer a powerful alternative for highly available applications.

Moreover, cloud paradigm could be attractive as it can improve the development of applications specifically designed to take advantage of cloud infrastructure. Meanwhile, new functionality could be provided, usage of cloud-based services can be scaled up and down smoothly without the need for upfront cost. Finally, cloud technology helps break the barriers between different web applications and reduces the fragmentation of data.

Beside the new possibilities offered by Cloud computing two significant concerns must be addressed before moving data and applications to a cloud environment: the security of information and the bottlenecks of data transferring.

*Information and privacy.*

The scalability and the easy access to cloud resources increase the risk of security of data that have security requirements out of the ordinary. Hosting such data on publicly accessible servers requires the definition of security rules to protect data security and privacy. For example, mechanisms such as encryption measures and role-based access control models should be deployed on the Cloud respecting both the typicality of data and the requirements of users in sharing such a data [89]. Differently from conventional web-based environments, the use of the Cloud for storing data and applications improves physical security as it enables the creation of separate virtual machines and firewalls for each independent application environment. Security professionals traditionally recommend partitioning a system as a means of protection. The ease of creating new virtual machines provides ways to improve the security of hosted data by partitioning resources into separate areas, conferring protections against attacks. Moreover, both data and server backups can be arranged easily and a security staff and tools are available by the Cloud provider.

*Bottlenecks of data transferring.*

Cloud vendors store backups of users' applications and data in multiple geographical locations. Moreover, cloud computing allows the sharing of data in real-time collaboration with other users. Although, a cloud can provide extra processing resources during the peaks (within limits), the transfer of vast amounts of data to the cloud is a significant bottleneck in cloud computing [90]. Networking bandwidth limitation causes delays in data transfer and incurs high bandwidth costs from service providers. This cost is an important issue for institutions that require substantial data movement (on the order of terabytes and petabytes) on a regular basis. As a consequence, the use of a cloud infrastructure currently does not make economic sense for applications that need to continuously export or import large volumes of data to and from the cloud.

## 3.2 Motivation for a New Architectural Framework

The increasing number of Open Data portals result in the provision of a large amount of data without any significant effort for improving their search. Regardless of the numerous initiatives, OGD portals have a limited scope as they tackle the ability of the governments to present their data to the public. So, they continue to offer a lot of datasets of a summary nature and pertaining to a plethora of activities.

Although the use of advanced open data platforms for the creation of OGD portals (such as data.gov.uk, data.gov.fr, digitaliser.dk, etc.), a gap can be perceived between the abundance of data and the lack of efficient and effective tools that enable data searching. Indeed, open data platforms provide simple tools for enriching governmental portals without new innovative approaches to facilitate the process between data consumers and data providers. Even though state-of-art open data platforms share common aims and objectives, however they differ in architecture, file formats, features and functions [91]. The lack of user-friendly tools for searching and integrating data makes difficult to formulate queries and capturing information requires a level of user expertise [92].

As a consequence, despite the high expectations, open data initiatives are gathering very limited interest in citizens due to many barriers including the complexity of searching specific information. Therefore, a need for new user-friendly tools arises, which enables citizens to search and capture government data. In this regard, the main open questions are:

- What kind of open government applications have the potential to favour a citizen-centric approach for understanding citizen demands about open data search?

- How to design, develop and implement such applications by selecting and combining the most advanced technologies to meet the requirements of the users?

The exponential growth of government data as well the availability of open data from several domains provides opportunities for carrying out new research about how harnessing data to deal with fundamental civic and government problems such as the organization of public services, urban planning, street construction planning, health emergency response planning, etc.

As well, the rapid development of the Internet has provided an opportunity to use state-of-the-art technology for facing the solution of such a problems by a new generation of ICT tools that integrate plain data sources, public programmable APIs and any kind of available services. Usually referred to as Web 2.0 applications, these tools rely on contents that are available on the web in form of open APIs or reusable services. Internet dramatically increases the range of benefits and usages derived from these applications that clearly point toward a high user involvement.

Cloud computing is a significant component of the Web 2.0 world as it expresses a service-based architecture where computing services are delivered on-demand to customers over a network in a self-service fashion that is independent of device and location. In particular, the PaaS level enables developer to use the cloud platform for developing applications that are offered as cloud services. Indeed, the developer of PaaS applications uses the software environment offered by the cloud platform (including programming languages, database management systems, tools, etc.) without dealing with software issues. The development tasks are facilitated and result in applications that scale up or down elastically according to their actual need for resources.

The application of cloud computing in managing OGD is yet preliminary [93] and few work has be done to explore the use of architectural cloud-based solutions that improve the functionality of open government portals [94]. Indeed, the adoption of a Cloud infrastructure poses important challenges concerning the design and implementation of applications that concentrate on introducing public value services for a web community that includes citizens, private stakeholders, civil servants and government bodies. So, their design requires efforts for adding context to data i.e. for making data ready-for-use by both government and the public. Context is closely related to the interests of various audiences. Although their effectiveness, government agencies concentrate their efforts in disclosing a lot of machine-readable data and fail to provide any context except the one that is relevant to the agency that owns the data. The other side of the coin is that it is unrealistic to think that agencies may predict all the application contexts or all the intended uses. However, it makes sense to investigate about how to harness state-of-art ICT technologies in order to favour the reuse of OGD for improving services for both the government and the external audience.

The re-usability of OGD depends on two basic factors i.e. the government effectiveness in providing open information and the ICT's advances over time. So government agencies and developers should be interested in improving each other by devising common strategies for organizing open data and making OGD ready-for-use for a large class of applications. As discussed in the previous chapter, momentum increases in governments and municipalities in publishing and promoting open data initiatives through the use of advanced platforms and OGD apps. These efforts should be consistent with similar initiatives by the research community to identify appropriate and advanced architectural environments to structure and reuse open data when OGD comes from government portals that rely on state-of-art technologies.

The intrinsic dynamism of open data suggests that research efforts should concentrate on devising architectural models for supporting the context of data as well suitable software environments for the development of value added OGD apps.

Accordingly, next section focuses on introducing a quite general architectural framework that relies on services offered at the PaaS level for developing a scalable catalogue that loosely organizes metadata captured from government platforms and other Internet resources. The catalogue supports the context of data and is a reference data model for implementing cloud-based applications that enable easy searching and downloading OGD from portals and their reuse for external initiatives. As well, the proposed framework promotes a new scenario enabled by flexible and easy-of-use services available in Internet, such as GIS services for accessing, structuring and analysing spatial data.

## 3.3 The Architectural Framework

The proposed framework aims to harness an integrated set of cloud-based services to:

- lower the complexity of integrating OGD regardless of its format and location;
- enable query and navigation support with high level of flexibility to the user needs;
- facilitate developers in the creation of OGD apps.

Its design has faced several practical challenges resulting from the following aspects. Open government portals are large in size, physically distributed, autonomously owned and operated. Consequently, there is the need for mechanisms that can efficiently extract, on demand, only the relevant information . The effectiveness of such mechanisms is bound by the quality of metadata that characterize data, which is not always good.

Being heterogeneous in structure and content, government portals organize metadata according to their own catalogues that, implicitly or explicitly, define concepts about data and relationships among concepts. These issues make hard the capture of metadata by automatic programs because the harvesting strategies must comply with the operational constraints imposed by the data source.

Although an important first step has be done by making open data and cloud computing part of the government policy, there is still a long road ahead towards enabling government agencies to deliver value to the public and private sectors in new and innovative ways.

The proposed framework is structured around cloud-based technologies and suitable services that:

- Support users in searching and accessing OGD, including browse, visualize, analyse & download OGD in multiple formats by several web sites.
- Gather data via open standards Application Programming Interfaces (APIs).
- Streamline and enhance the development of innovative applications from developers.
- Use cloud services for lowering up-front infrastructure costs (servers, software, etc.).
- Allow data requirements to be met in an on-demand and scalable manner via cloud.
- Guarantee reliability and scalability when the volume of data increases compute requirements.
- Consume on the Internet multiple catalogues of public data sets in a way that encompasses and meets the fundamental principles of the Web 2.0.

Figure 11 shows the main components of the proposed architecture: a catalogue of metadata, a procedural component and a user interface.

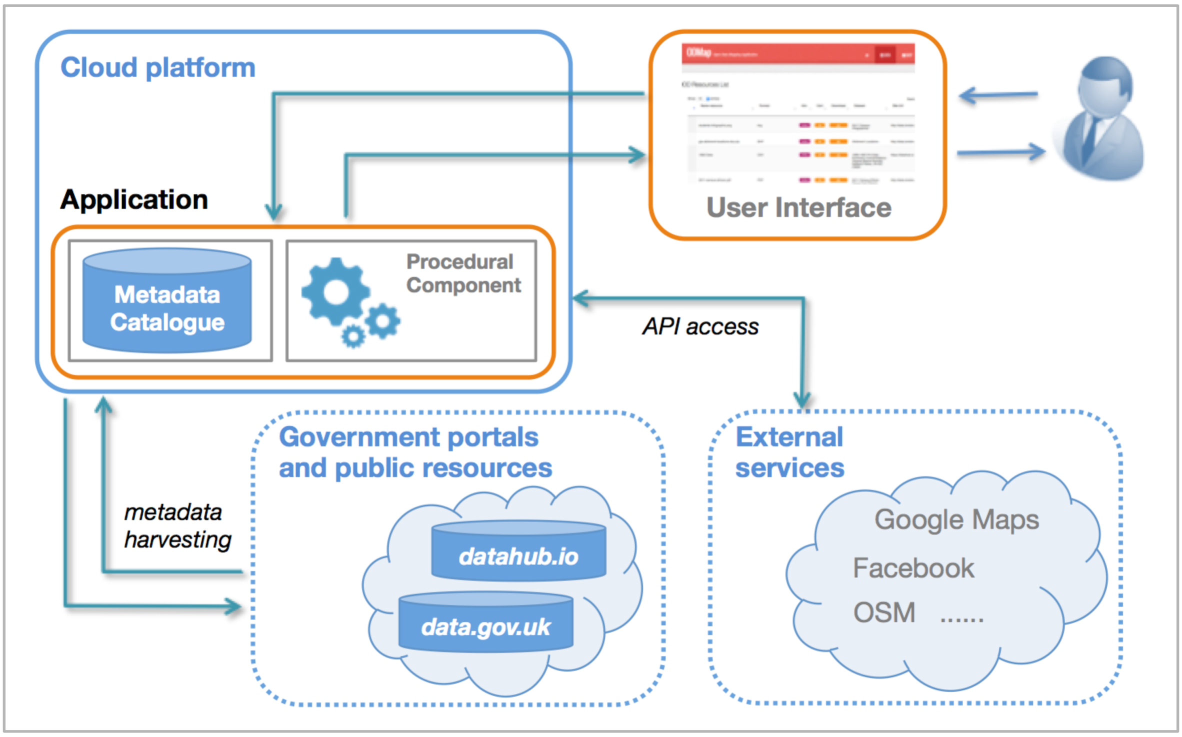



Figure 11 - The Proposed Architecture.

The **Catalogue** is the core of the proposed architecture and relies on metadata harvested from distributed open government portals. It enables the context of OGD by a multi-level index that clusters harvested metadata into sub-catalogues whose content depends on the particular tasks of the application. The Catalogue is enabled by the NoSQL database offered by cloud environment. This gives a great flexibility in organizing metadata and avoids the definition of an a priori schema for matching distributed information. Moreover, the use of a DBMS as a cloud service improves the data management in terms of elasticity and scalability. Indeed, the catalogue is not necessarily stored on a single physical device (a hard drive or a file system) but across many virtual machines, allowing to maintain high performance when the catalogue enlarges for the purpose of enabling search over additional sites.

The **procedural component** addresses challenges associated with the development of OGD apps over several distributed resources. It supports discovering, aggregating and delivering resources and services according to user's goals, tasks and concerns and provides the following types of services:

*Data Management Services*

These services enable operations concerning the access to government portals and distributed web resources by tracking portals and remote data sources so that users can improve their agility in downloading OGD without worry about technical details such as file formats, input/output parameters, etc. Data Management Services rely on Uniform Resource Identifiers (URIs) for resource identification. As well they capture data for programmatic access by means of web APIs made available by the government sites and external providers. This allows developers to start quickly on new applications using a variety of client technologies (e.g. JavaScript, Python, PHP, etc.) for accessing APIs.

These services exploit a number of formats (e.g. Open Data Protocol (OData), JSON, Keyhole Markup Language (KML), etc.) for enhancing data integration and interoperability across a broad range of clients, servers, services, protocols and tools and returns geospatial data in standard formats (e.g. the KML, geoJSON, etc.) that are compatible with popular desktop and web-based mapping technologies including Microsoft Bing Maps, Google Maps, Yahoo! Maps, OSM, and Google Earth. Users and developers can quickly download data in their own computers and use them for developing applications.

*Task-oriented services*

These services are implemented by developers of applications that exploit the context of data defined by the catalogue. Mash-ups are a flexible way for their customized composition.

*Data Administration Services*

The data administration deals with the accommodation of information in the catalogue, supports the developers in quickly loading the data for populating the catalogue, create new datasets, add data, or update data in the catalogue.

The **user interface** uses the context defined by the catalogue and the task-oriented services to provide effective support to OGD search and management. It encompasses two interaction mode (i.e. data search and data administration) and exploits visual tools for user-friendly data searching. OGD and other data are combined and visualized in widely used and recognizable formats such as tables and maps. Visualization, mashups of data, and map projection help handle information overload. When required, users are allowed to formulate spatial queries to obtain significant and precise information in spaces of the real world.

The proposed framework is quite general and served as a reference for the development of the two applications presented in next sections. Indeed both applications have been implemented on a PaaS cloud platform and benefit from web technologies to offer services to users with different needs and requirements. As well, a wide range of applications can be implemented starting from the framework that aggregate content from distributed OGD resources to provide services to end-users with different interests and skills.

Beyond the simple division between not experienced (i.e. the ordinary citizen) and experts (i.e. professionals, researcher or journalist) users, is useful to consider the their interests and commitment of other categories of users including the evaluator, the beginner, the intermediate user, the expert user and the power user [88].

The goal of the evaluator is to investigate the application, understand how it works and what it can do. The application should give a good first impression, the design and implementation of the user interface plays an important role and, at a first level, must not be unnecessarily complicated. It is to provide useful how-to videos, tutorials with step-by-step instructions and, if necessary, documentation and sample files to experiment with the application.

For beginners who try to use the application by exploring and gradually also learning from mistakes, it is important that the application offers an easy identification of features and their use. This can be done applying a good visual design, information architecture and interaction design. The provision of wizards to configure a project and the ability of undo/redo encourage experimentation and remedy potential errors. Also getting started guides have to be supplied, accompanied by manuals, in-depth tutorials, online help and context.

The intermediate user is usually characterized by the periodic use of the application in order to meet targeted needs, and usually is not interested in all the functionality. This type of user certainly will use indexed and searchable online help.

Advanced users are generally characterized by a strong knowledge of the application domain and are able to quickly complete a complex task. For this type of user should be possible to bypass the wizards and turn off any pop-ups or other types of aid intended for the beginner users.

The power users are very interested in the application and try to exploit it to the maximum. If allowed, they are able to customize and automate certain tasks. This type of user can take advantage from forums and from the exchange of tips with other users.

In addition to the above, to train users to a profitable use of the application, it is desirable to integrate a system of collaboration with the ability to report any malfunctions and useful feedback to administrators and developers.

Next chapters present two case studies that validate the proposed framework over GAE (Google App Engine) [14], a PaaS environment, which enables the development and the deployment of cloud-based applications. Differently from other PaaS offerings, GAE benefits from the same infrastructure that supports Google applications and services such as Google search engine, YouTube, Google Earth, etc. In particular, GAE offers the Datastore (namely Cloud Datastore), a distributed data storage service that performs distribution, replication and load balancing automatically. The Datastore provides many capabilities such as ACID transactions and supports operations to access data objects (i.e. create, read, update, delete) and indexes. Queries are also allowed by means of an SQL-like language called GQL. Specifically, the Datastore is an object-oriented and schemaless NoSQL database [95] [97] that stores data objects clustered in classes denoted as "kinds". Each data object has a key (that uniquely identifies it within the Datastore) and is featured by properties belonging to several data types (e.g. string, integer, URL, etc.). Key values are used for referencing data objects among them as it happens in relational databases to define the referential integrity between two relations. Table 9 shows the analogy between common terms used in object oriented and relational database, and Datastore terms.

| Object-Oriented | Relational Database | Datastore |
|---|---|---|
| Class | Table | Kind |
| Object | Record | Entity |
| Attribute | Column | Property |

**Table 9 - Analogy between terms in Object-Oriented, Relational Database and Datastore.**

Notable, the Datastore provides a RESTful interface, data can easily be accessed by any deployment target. It is possible to build applications that utilize others Google storage resources (e.g. Cloud SQL), and rely on Datastore as the integration point.

# 4 The first case study: Open Data Mapping (ODMap)

As previous mentioned, tools made available by government portals often offer limited functionality to search OGD. Moreover, it is common to find websites that offer no support to filter the datasets in accordance with the user's needs. Conversely, several portals support search facilities that return a lot of content including not only the data of interest but also related policies and government documents such as technical reports and supplementary maps. These drawbacks make worse the problem of locating relevant data for users. As well, developers have difficulty in organizing in a single context, the amount of data from different sites.

ODMap (Open Data Mapping) is a prototypical cloud-based application that emphasizes the interests of citizens in accessing data belong to a specific context by defining dynamic patterns of interaction that increase the interest of the users in exploring OGD [102]. Specifically, ODMap undertakes the responsibility of supporting user in searching over tens of decentralized CKAN portals managed by different organizations.

A complementary case study [96] sets out the application of the proposed framework on SIOPE, an OGD Italian portal that provides access to the receipts and payments made by the Italian government institutions [97].

As showed in Figure 12, ODMap harvests metadata from portals and organize them into a flexible catalogue that helps users in identifying the relevant OGD.

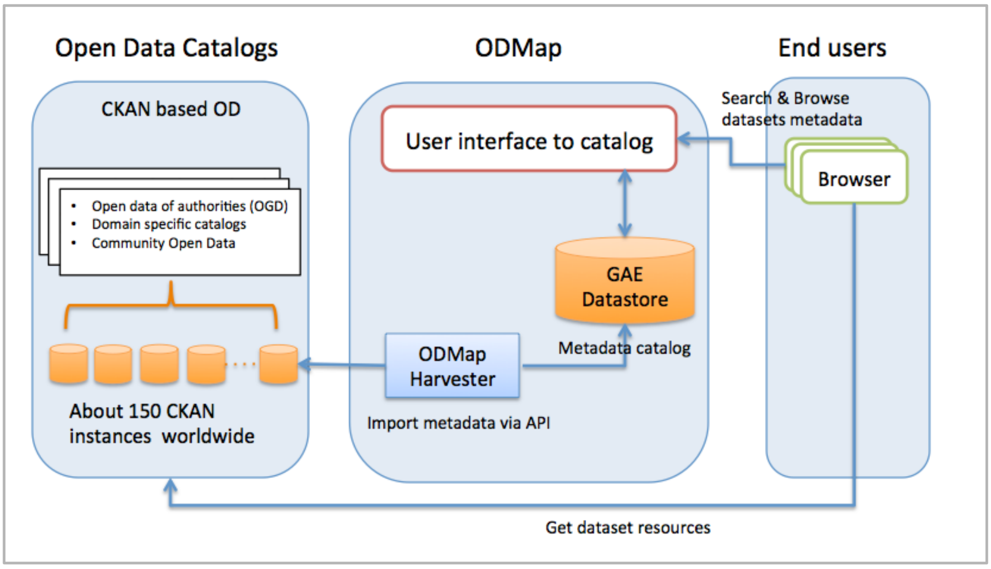



Figure 12 - ODMap basic functionality.

Specifically, the ODMap Catalogue clusters metadata harvested from the CKAN sites into the following sub-catalogues:

▪ *ODSites* - includes metadata (such as the URL, the name, the region, etc.) about the considered CKAN sites.

▪ *Datasets* - contains metadata about datasets stored into CKAN repositories such as URL, title, name, resources, etc.

▪ *Tags* - stores tags attached to datasets.

▪ *Organizations* - stores information about the organizations that own and manage datasets and registered users. Note that only administrator users are allowed to manage (insert/update/cancel) datasets.

Each sub-catalogue is stored in a specific kind of the GAE Datastore. Figure 13 and Figure 14 show the definitions of kinds for the sub-catalogue ODSites and the sub-catalogue Dataset.

```
class ODSites(db.Model):
    id = db.StringProperty(required=True)
    typesite = db.StringProperty()       # CKAN, DKAN, ...
    region = db.StringProperty()         # Europe, Asia, North America ...
    url = db.LinkProperty(required=True)
    urlapi = db.LinkProperty()
    name = db.StringProperty(required=True)
    description = db.StringProperty()
    country = db.StringProperty()
    city = db.StringProperty()
    location=db.GeoPtProperty()  # by latitude, then longitude es.(39.251,9.173)
    point=db.TextProperty()      # geojson or kml
    bound=db.TextProperty()      # geojson or kml
```

**Figure 13 - Definition of ODSites kind.**

```
class Dataset(db.Model):
     id = db.StringProperty(required=True)
     title = db.StringProperty()
     name = db.StringProperty(required=True)
     url = db.LinkProperty(required=True)
     notes = db.TextProperty()
     num_tags = db.IntegerProperty()
     num_resources = db.IntegerProperty()
     tags = db.TextProperty()
     resources = db.TextProperty()
     organization = db.TextProperty()
     extras = db.TextProperty()            # key-value spatial data
     package_full = db.TextProperty()      # all dataset metadata
     geo_box = db.StringProperty()         # geojson type=polygon
     geo_point = db.GeoPtProperty()        # by latitude, then longitude
```

**Figure 14 - Definition of Dataset kind.**

The following additional kinds express the relationship among the sub-catalogues:

- *DatasetTag* links tags to the datasets they feature.
- *DatasetOrganization* links datasets to the organizations that manage them.
- *DatasetResource* relates datasets to the resourcesthat compose them.

The procedural component addresses challenges associated with searching data over several distributed resources. Data is captured using APIs made available by the CKAN sites [99]. CKAN's API is a powerful, RPC-style API that exposes all core features to API clients. The Data service exploits the API to get a full JSON representation of datasets, resources or other objects. Specifically, data from datastore are parsed in JSON to be rendered to the user interface.

OGD including geospatial data is returned in GeoJSON format, making it compatible with popular desktop and Web-based mapping technologies including Microsoft Bing Maps, Google Maps, Yahoo! Maps, OSM, and Google Earth. JavaScript Open source libraries enable the creation of dynamic maps such as OpenLayers [100] e Leaflet [101].

Administrative tasks mainly focus on harvesting metadata and populate the catalogues according the following steps. First, massive harvesting is automatically performed to populate the sub-catalogue “ODSites” which currently hosts metadata about tens of CKAN based sites. Then, metadata are extracted about the datasets stored in those sites and their content to populate the sub-catalogue “Datasets”. Finally, the sub-catalogue “Tags” and the sub-catalogue “Organizations” are populated by tags and metadata about organizations. All the sub-catalogues can be automatically updated by adding single instances e.g. an additional CKAN site, a single dataset, etc.

Figure 15 shows an example of tags extraction and indexing. From the property “tags” of each entity in *Datasets* are extracted all tags related to each dataset and stored in an indexed kind named *DatasetTag.*

Geographic data are very important because they allow users to visualize and select the sites of interest. Figure 16 shows an example of such visualization. Here the user obtains detailed information about the locations of the 151 available CKAN sites stored in ODMap. Then, he selects in the map an Indonesian CKAN site and is provided with specific information about this site.

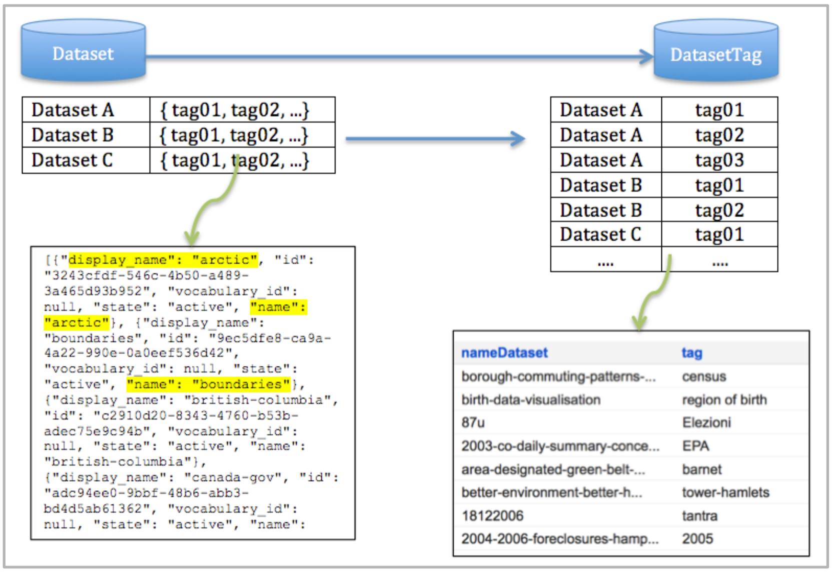



Figure 15 - ODMap Metadata indexing: from JSON to key-value like kind.

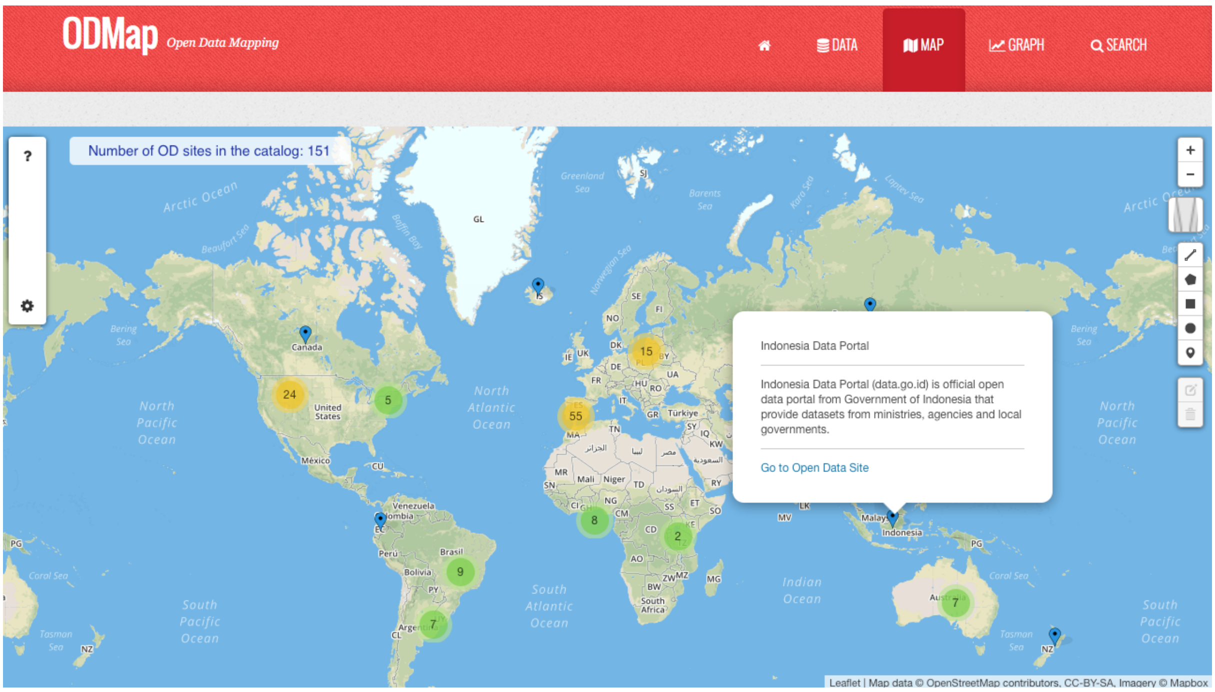



Figure 16 - OGD Sites visualized on worldwide map.

When a geo-reference is not automatically available, a manual geo-reference can be stored as showed in Figure 17.

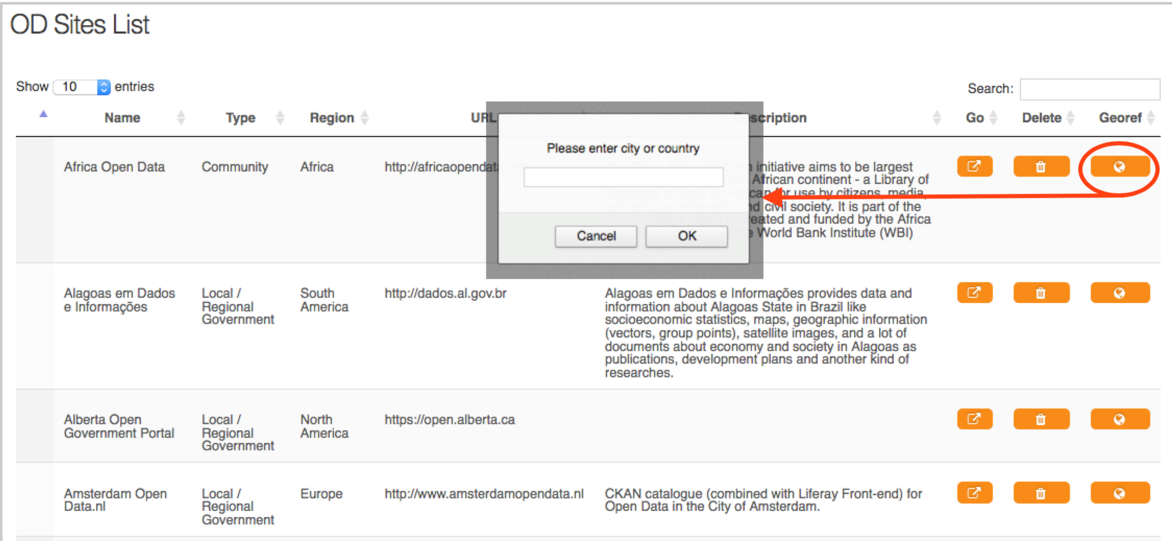



**Figure 17 - Manual geo-referencing of a site.**

ODMap aims to offer new ways of searching OGD using some contexts of data that support tasks about common points of interest the users could share.

Figure 18 and Figure 19 show with informal notation, respectively, the ODMap context diagram (describing the relationships, dependencies, and interactions between the system and its environment) and the containers' diagram that describes system components such a web servers, application servers, mobile apps, databases, file systems, desktop applications, etc. Essentially, a container is any component that can host programming code or data.

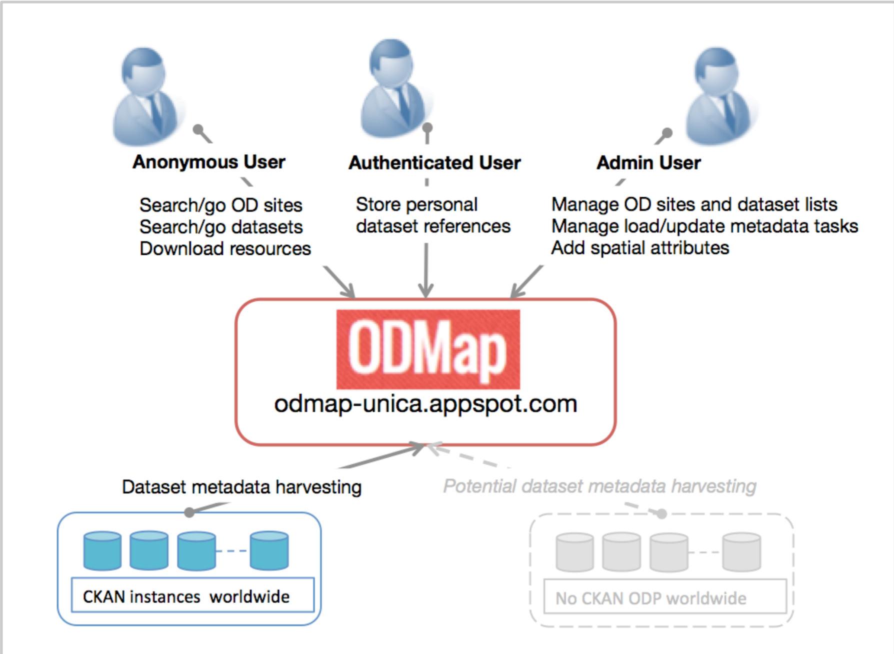



Figure 18 - ODMap Context diagram.

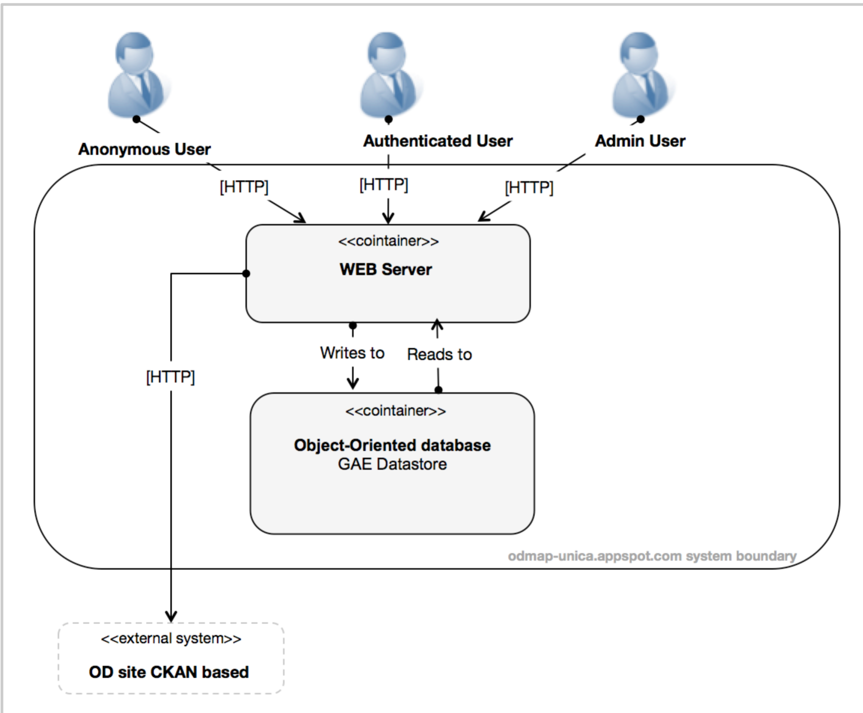



Figure 19 - ODMap Containers' diagram.

Figure 20 shows the ODMap user interface and his main contexts and the Table 10 summarizes the functionality of various buttons.

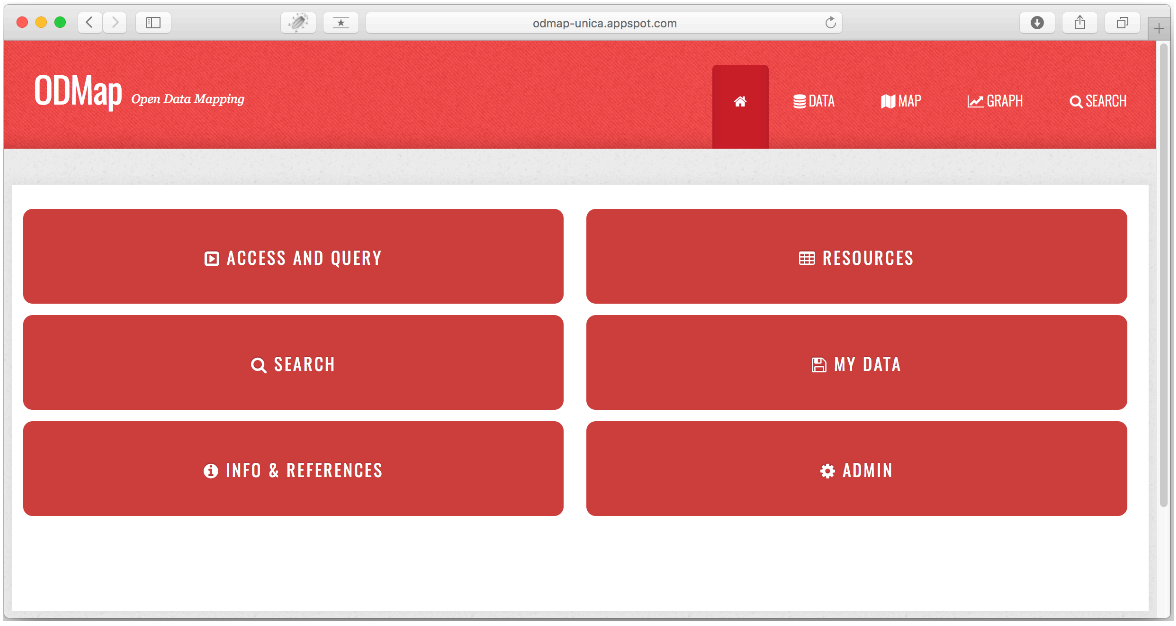



Figure 20 - ODMap Home Page.

| Menu item | Description |
|---|---|
| Access and Query | It allows users to access OGD portals managed by ODMap. |
| Search | It allows to query dataset metadata:<br>- by tag<br>- by spatial search |
| Resources | It allows the resources selection by:<br>- open data sites<br>- dataset<br>- tags<br>- organizations<br>- resources (files in various formats) |
| My Data | It allows to access a personal repository of metadata and the recovery of related datasets. |
| Info & References | It allows to access the information section of the application. |
| Admin | It supports the access to administrative functions including the harvesting of metadata and their management. |

Table 10 - ODMap Main menu description.

Specifically, the user interface in Figure 20 supports the following functionalities.

*ACCESS AND QUERY* – It allows users to search datasets by specifying tags (i.e. keywords) or locations over a map (see Figure 21).

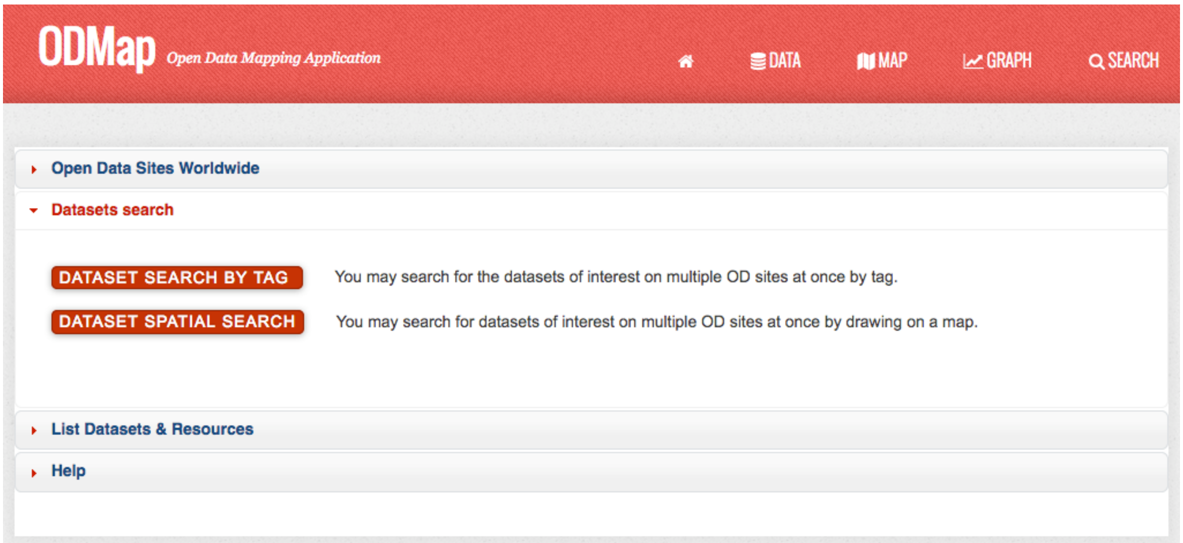
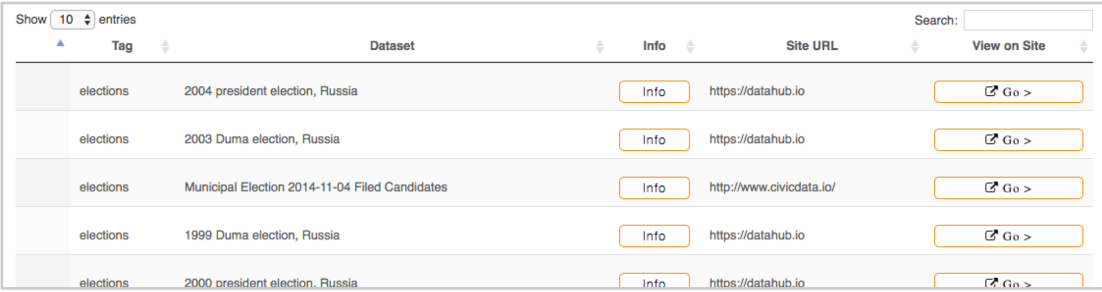



Figure 21 - The Access and Query page.

*SEARCH* – It enables search over tags and spatial queries. As Figure 22 shows, ODMap visualizes a cloud of available tags using fonts whose magnitude is proportional to tag's frequency in the Datastore. Figure 23 shows the result of a search.


Figure 22 - Search Datasets by tag.



Show 10 entries Search:

| Tag | Dataset | Info | Site URL | View on Site |
|---|---|---|---|---|
| elections | 2004 president election, Russia | Info | https://datahub.io | Go > |
| elections | 2003 Duma election, Russia | Info | https://datahub.io | Go > |
| elections | Municipal Election 2014-11-04 Filed Candidates | Info | http://www.civicdata.io/ | Go > |
| elections | 1999 Duma election, Russia | Info | https://datahub.io | Go > |
| elections | 2000 president election, Russia | Info | https://datahub.io | Go > |

Figure 23 - Example of datasets retrieved by specifying the tag "elections".

Figure 24 shows an example of spatial search. Here, the user draws the rectangular red box on the map. Correspondingly, ODMap retrieves all the datasets geo-referenced within such box (see Figure 25) over all previously collected CKAN sites, as it happens searching by tag.

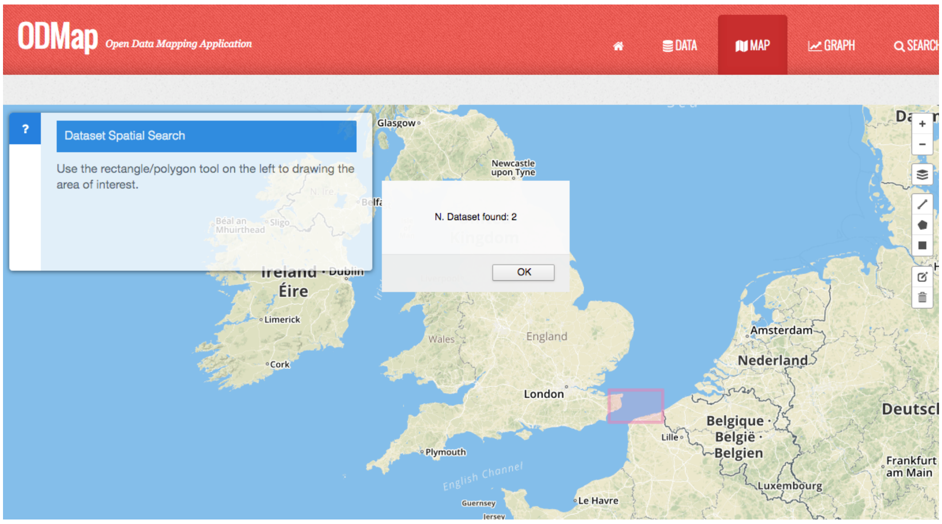
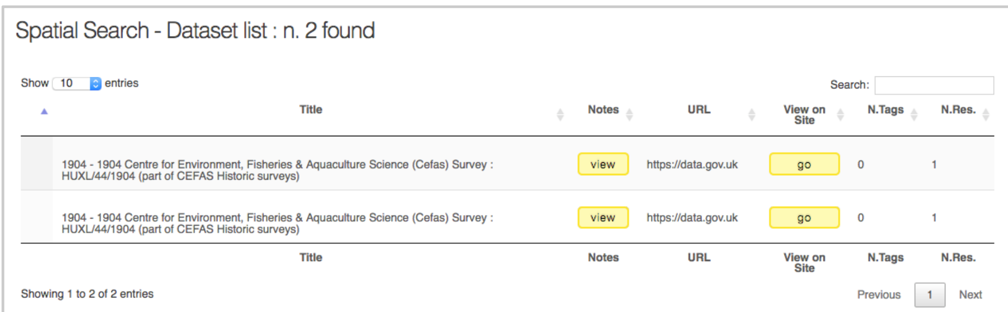



Figure 24 - ODMap Spatial Search.



Spatial Search - Dataset list : n. 2 found

Show 10 entries Search:

| Title | Notes | URL | View on Site | N.Tags | N.Res. |
|---|---|---|---|---|---|
| 1904 - 1904 Centre for Environment, Fisheries & Aquaculture Science (Cefas) Survey : HUXL/44/1904 (part of CEFAS Historic surveys) | view | https://data.gov.uk | go | 0 | 1 |
| 1904 - 1904 Centre for Environment, Fisheries & Aquaculture Science (Cefas) Survey : HUXL/44/1904 (part of CEFAS Historic surveys) | view | https://data.gov.uk | go | 0 | 1 |
| Title | Notes | URL | View on Site | N.Tags | N.Res. |

Showing 1 to 2 of 2 entries Previous 1 Next

Figure 25 - Example of datasets retrieved by spatial search.

*RESOURCES* – This task allows to browse, filter, search and select available resources (i.e. open data sites, datasets, organizations, etc.) according to filtering criteria. Figure 26 shows the options offered by this task. Table 11 describes in detail the supported functionalities.

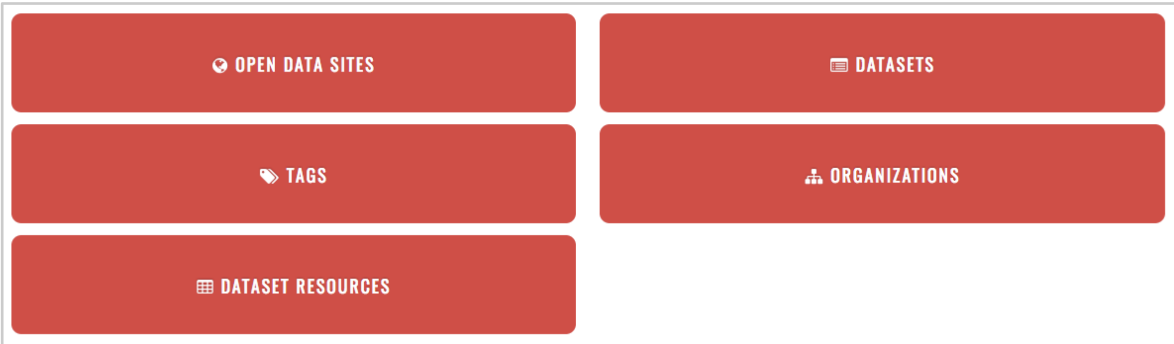
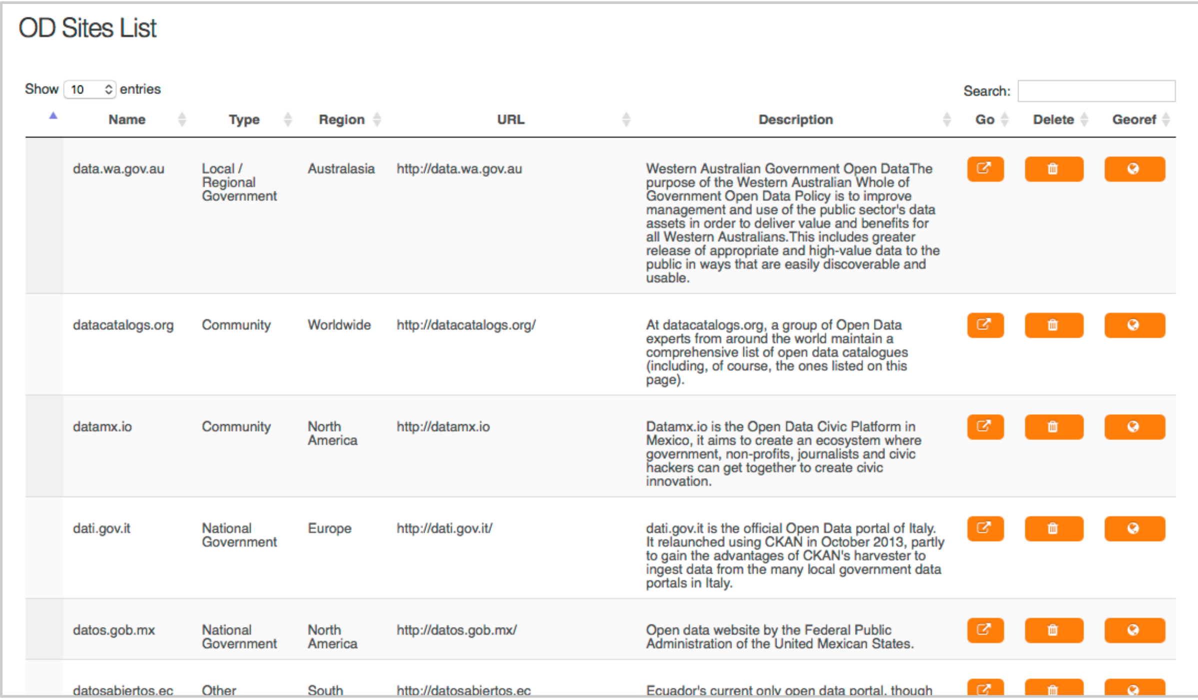



**Figure 26 - The RESOURCE menu.**

| Menu items | Functions | Figures |
| --- | --- | --- |
| Open Data Sites | Lists the CKAN web sites from which metadata have been or could been harvested. | Figure 27 |
| Datasets | Lists datasets whose metadata is stored by ODMap. | Figure 28 |
| Tags | Lists tags of the datasets whose metadata is stored by ODMap. | Figure 29 |
| Organizations | Lists the organizations that have published the datasets whose metadata is stored by ODMap. | Figure 30 |
| Dataset Resources | Lists files that make up the datasets whose metadata is stored by ODMap. | Figure 31 |

**Table 11 - The RESOURCE functionalities.**

*Open Data Sites* – This option presents the list of sites from which metadata can be harvested (Figure 27).

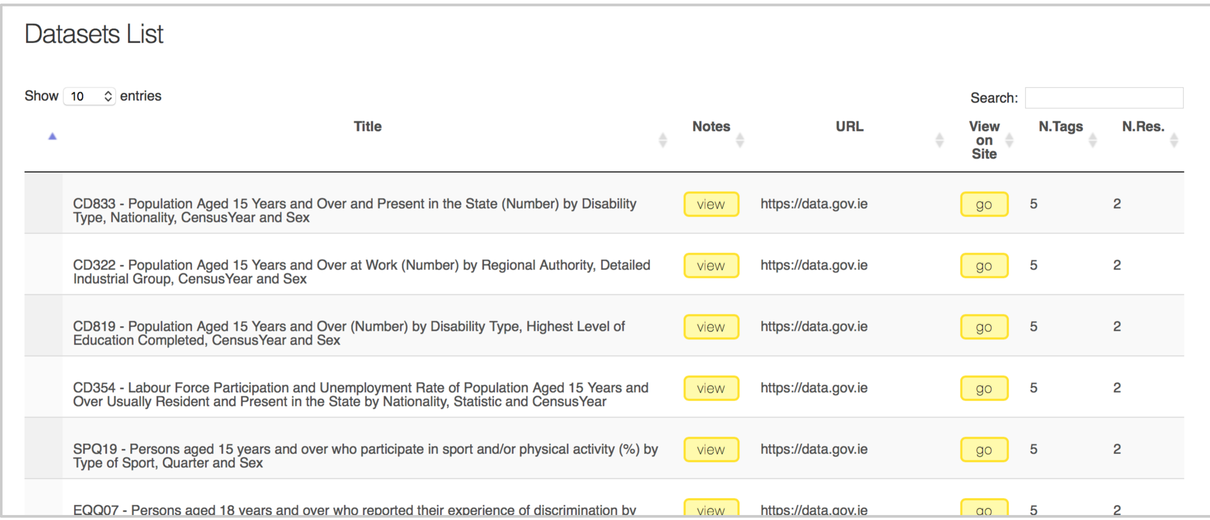

OD Sites List

Show 10 entries | Search:

| Name | Type | Region | URL | Description | Go | Delete | Georef |
|---|---|---|---|---|---|---|---|
| data.wa.gov.au | Local / Regional Government | Australasia | http://data.wa.gov.au | Western Australian Government Open DataThe purpose of the Western Australian Whole of Government Open Data Policy is to improve management and use of the public sector's data assets in order to deliver value and benefits for all Western Australians.This includes greater release of appropriate and high-value data to the public in ways that are easily discoverable and usable. | | | |
| datacatalogs.org | Community | Worldwide | http://datacatalogs.org/ | At datacatalogs.org, a group of Open Data experts from around the world maintain a comprehensive list of open data catalogues (including, of course, the ones listed on this page). | | | |
| datamx.io | Community | North America | http://datamx.io | Datamx.io is the Open Data Civic Platform in Mexico, it aims to create an ecosystem where government, non-profits, journalists and civic hackers can get together to create civic innovation. | | | |
| dati.gov.it | National Government | Europe | http://dati.gov.it/ | dati.gov.it is the official Open Data portal of Italy. It relaunched using CKAN in October 2013, partly to gain the advantages of CKAN's harvester to ingest data from the many local government data portals in Italy. | | | |
| datos.gob.mx | National Government | North America | http://datos.gob.mx/ | Open data website by the Federal Public Administration of the United Mexican States. | | | |
| datosabiertos.ec | Other | South | http://datosabiertos.ec | Ecuador's current only open data portal, though | | | |

Figure 27 - OD Sites list.

*Datasets* – This option enables listing all data sets whose metadata is stored in ODMap (Figure 28).

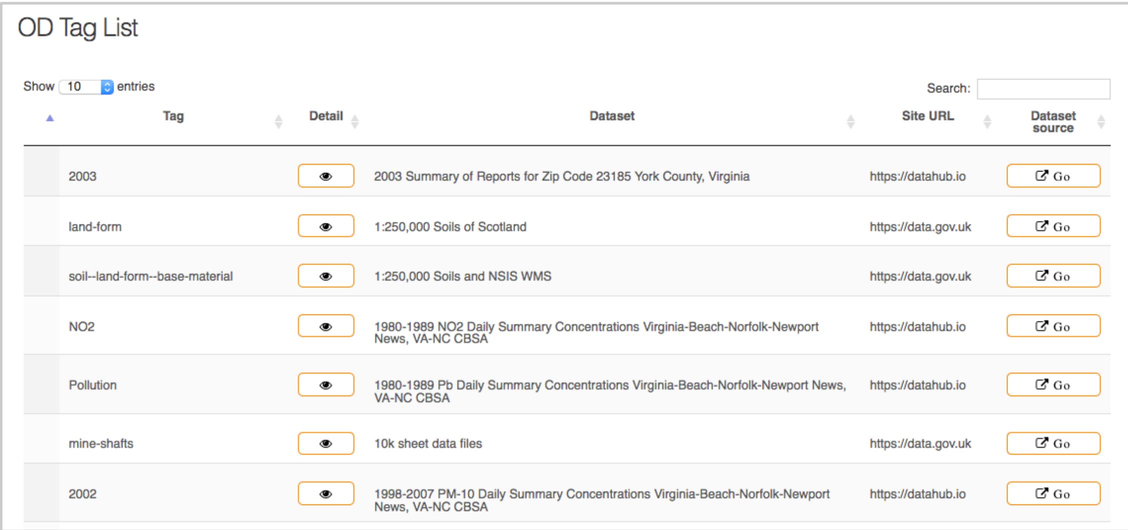

Datasets List

Show 10 entries | Search:

| Title | Notes | URL | View on Site | N.Tags | N.Res. |
|---|---|---|---|---|---|
| CD833 - Population Aged 15 Years and Over and Present in the State (Number) by Disability Type, Nationality, CensusYear and Sex | view | https://data.gov.ie | go | 5 | 2 |
| CD322 - Population Aged 15 Years and Over at Work (Number) by Regional Authority, Detailed Industrial Group, CensusYear and Sex | view | https://data.gov.ie | go | 5 | 2 |
| CD819 - Population Aged 15 Years and Over (Number) by Disability Type, Highest Level of Education Completed, CensusYear and Sex | view | https://data.gov.ie | go | 5 | 2 |
| CD354 - Labour Force Participation and Unemployment Rate of Population Aged 15 Years and Over Usually Resident and Present in the State by Nationality, Statistic and CensusYear | view | https://data.gov.ie | go | 5 | 2 |
| SPQ19 - Persons aged 15 years and over who participate in sport and/or physical activity (%) by Type of Sport, Quarter and Sex | view | https://data.gov.ie | go | 5 | 2 |
| EQQ07 - Persons aged 18 years and over who reported their experience of discrimination by | view | https://data.gov.ie | go | 5 | 2 |

Figure 28 - Datasets list.

*Tags* – This option lists all the tags of the datasets whose metadata are stored in ODMap (Figure 29).

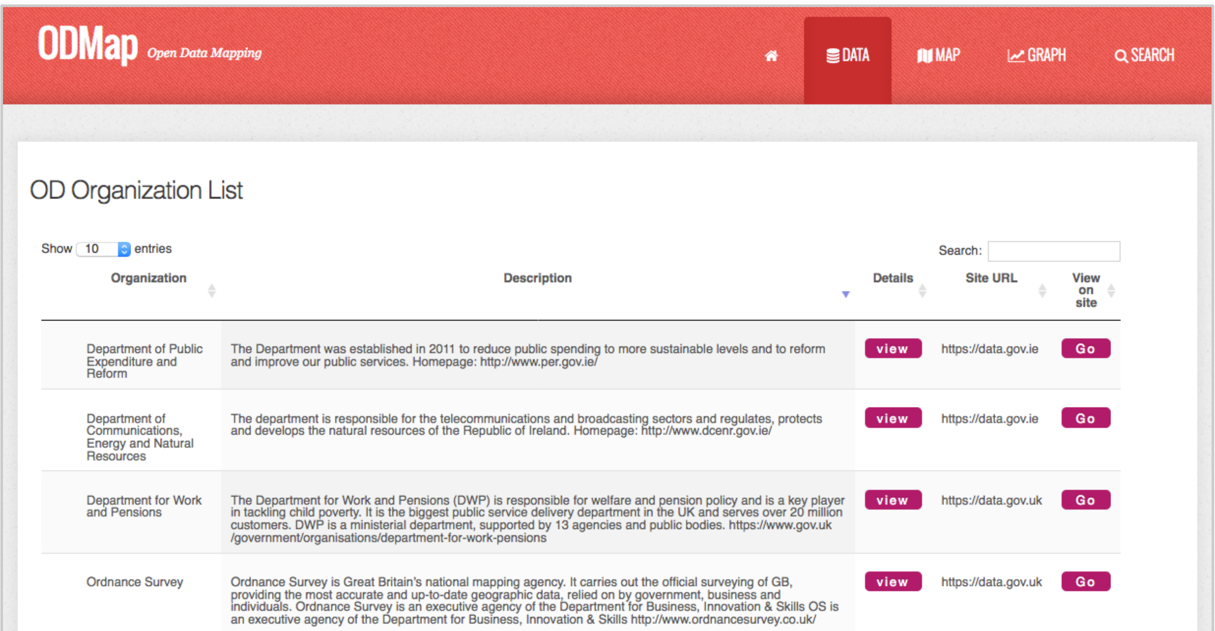

OD Tag List

Show 10 entries Search:

| Tag | Detail | Dataset | Site URL | Dataset source |
|---|---|---|---|---|
| 2003 | | 2003 Summary of Reports for Zip Code 23185 York County, Virginia | https://datahub.io | Go |
| land-form | | 1:250,000 Soils of Scotland | https://data.gov.uk | Go |
| soil--land-form--base-material | | 1:250,000 Soils and NSIS WMS | https://data.gov.uk | Go |
| NO2 | | 1980-1989 NO2 Daily Summary Concentrations Virginia-Beach-Norfolk-Newport News, VA-NC CBSA | https://datahub.io | Go |
| Pollution | | 1980-1989 Pb Daily Summary Concentrations Virginia-Beach-Norfolk-Newport News, VA-NC CBSA | https://datahub.io | Go |
| mine-shafts | | 10k sheet data files | https://data.gov.uk | Go |
| 2002 | | 1998-2007 PM-10 Daily Summary Concentrations Virginia-Beach-Norfolk-Newport News, VA-NC CBSA | https://datahub.io | Go |

Figure 29 - Tags list.

*Organizations* – This option lists all the organizations that have published the datasets whose metadata are stored in ODMap (Figure 30).



OD Organization List

Show 10 entries Search:

| Organization | Description | Details | Site URL | View on site |
|---|---|---|---|---|
| Department of Public Expenditure and Reform | The Department was established in 2011 to reduce public spending to more sustainable levels and to reform and improve our public services. Homepage: http://www.per.gov.ie/ | view | https://data.gov.ie | Go |
| Department of Communications, Energy and Natural Resources | The department is responsible for the telecommunications and broadcasting sectors and regulates, protects and develops the natural resources of the Republic of Ireland. Homepage: http://www.dcenr.gov.ie/ | view | https://data.gov.ie | Go |
| Department for Work and Pensions | The Department for Work and Pensions (DWP) is responsible for welfare and pension policy and is a key player in tackling child poverty. It is the biggest public service delivery department in the UK and serves over 20 million customers. DWP is a ministerial department, supported by 13 agencies and public bodies. https://www.gov.uk/government/organisations/department-for-work-pensions | view | https://data.gov.uk | Go |
| Ordnance Survey | Ordnance Survey is Great Britain's national mapping agency. It carries out the official surveying of GB, providing the most accurate and up-to-date geographic data, relied on by government, business and individuals. Ordnance Survey is an executive agency of the Department for Business, Innovation & Skills OS is an executive agency of the Department for Business, Innovation & Skills http://www.ordnancesurvey.co.uk/ | view | https://data.gov.uk | Go |

Figure 30 - Organizations list.

*Dataset Resources* –This option lists all the files that belong to the datasets whose metadata are stored on ODMap (Figure 31).

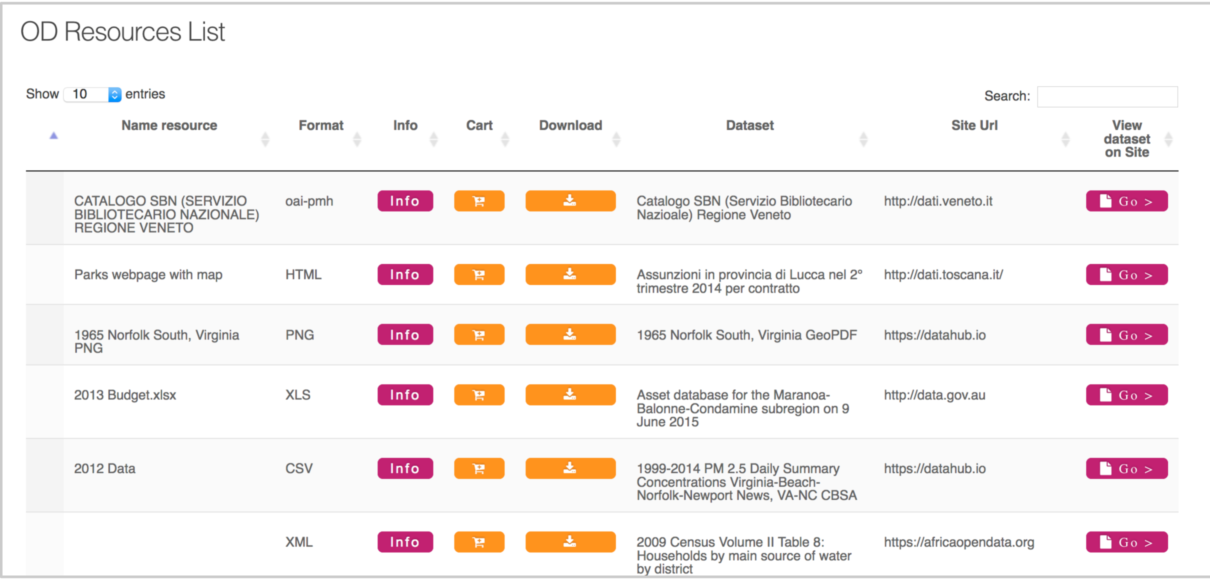



**Figure 31 - Dataset Resources list.**

*MY DATA* – It enables users to download data as well to create personal repositories of metadata (that annotate the resources i.e. datasets or files) and links to resources.

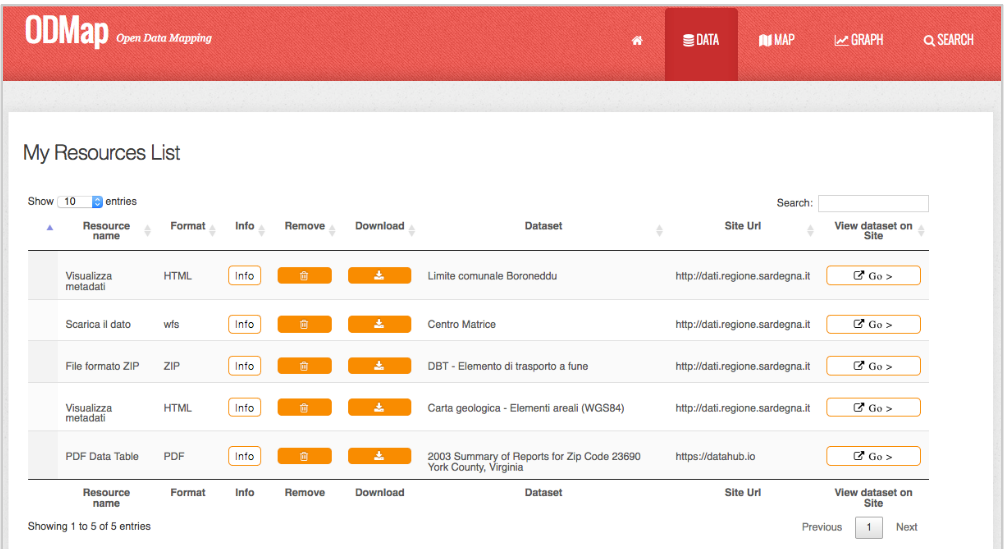



**Figure 32 - My Resources List.**

*INFO E REFERENCES* – As shown in Figure 33, this option presents general information about ODMap.

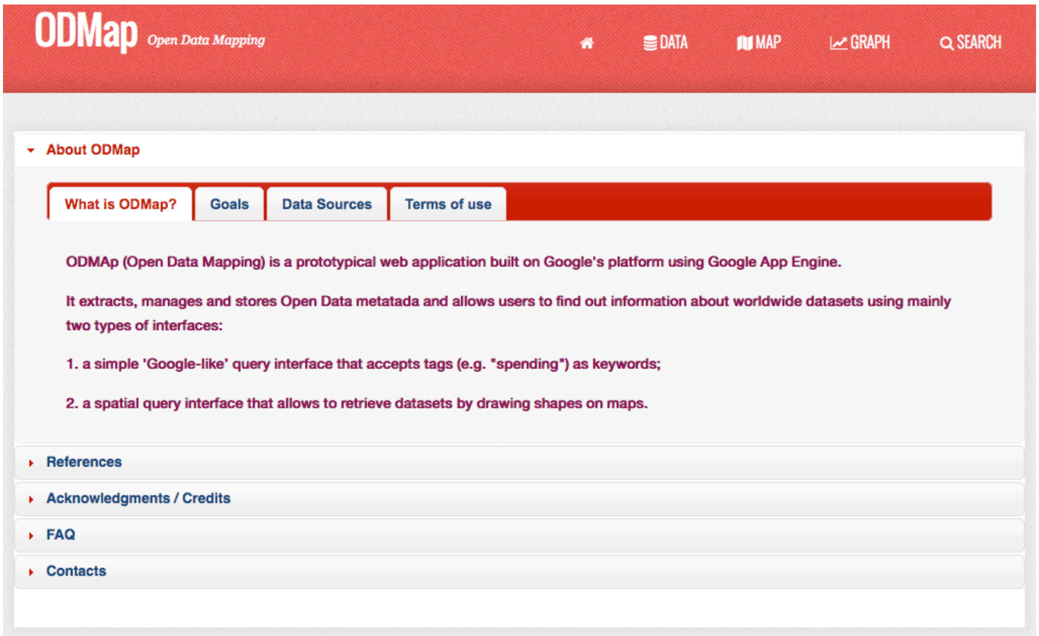



Figure 33 - Info & References.

*ADMIN* – It specifies the administrative functions of ODMap. An accordion menu groups available tools (Figure 34 and Figure 36) . The underlying Table 12 describes the various tools groups.

| Groups | Main tasks |
|---|---|
| Exploratory tools | Investigate the characteristics of open data sites and their metadata using APIs. |
| Harvesting tools | Evaluate the quantity and quality of metadata and import them into the ODMap's repository. |
| Manage dataset metadata | Organize metadata inside the repository. |
| Utility tools | General-purpose tools allowing to manually input or delete metadata and information about a site. |
| Help | Give some hints about how harvesting the metadata workflow and how managing it. |

Table 12 - ODMap: Groups of administrative tools.

The App Engine architecture is well suited for handling web requests, small amounts of work that run in a stateless environment with the intent of returning a response to the user as fast as possible. In our case, the extraction and storage of quite relevant amounts of data cannot be performed in the limit of 60 seconds imposed by the platform. After the evaluation step (Figure 35), the administrator can try to do the next action, e.g. to extract & store tags. If this action fails, he can create a new task that will be performed in background and/or in a deferred programmed time. To achieve this, ODMap code calls the task queue service to request a task. The task queue manages the process of driving the task to completion. By means of a specific console provided by App Engine platform (Figure 37), the administrator can manage and control the status of queues and tasks and take advantage of the detailed logs to troubleshoot errors.

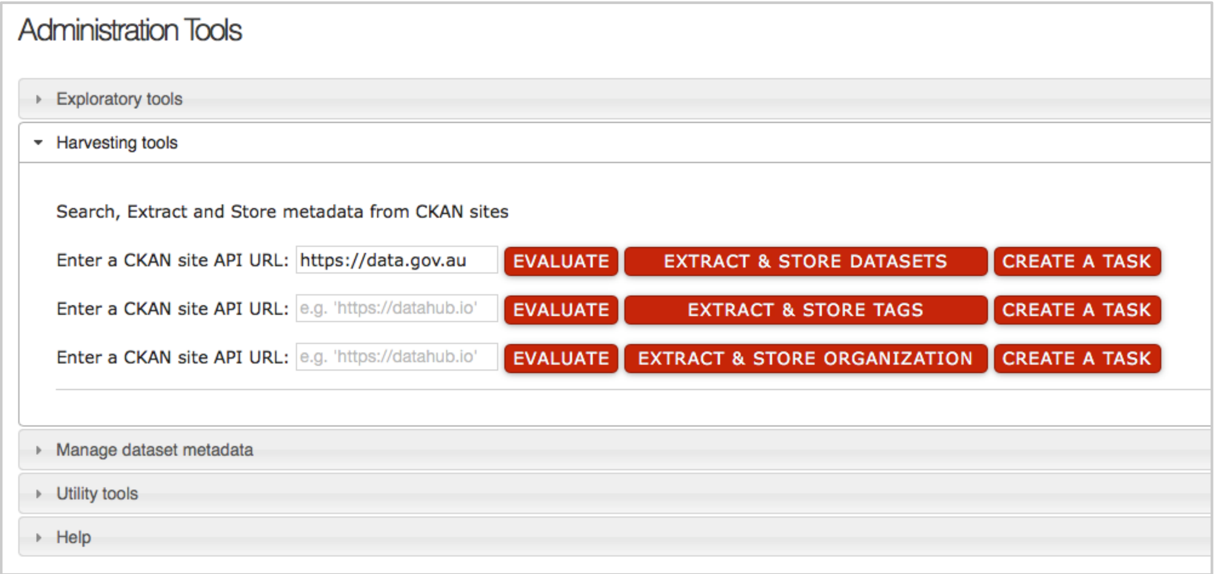



Figure 34 - Admin menu: Harvesting tools.

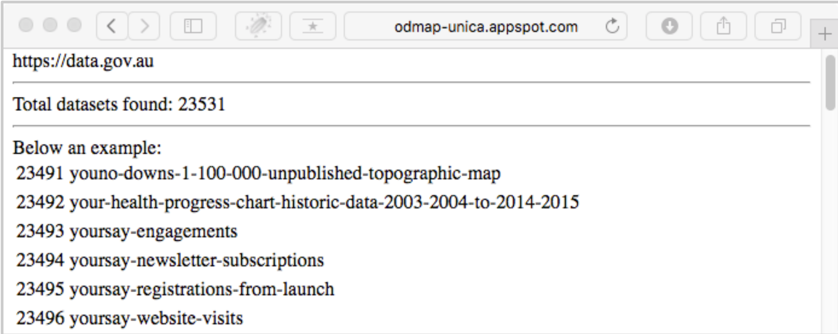



Figure 35 - Metadata harvesting: Evaluation step.

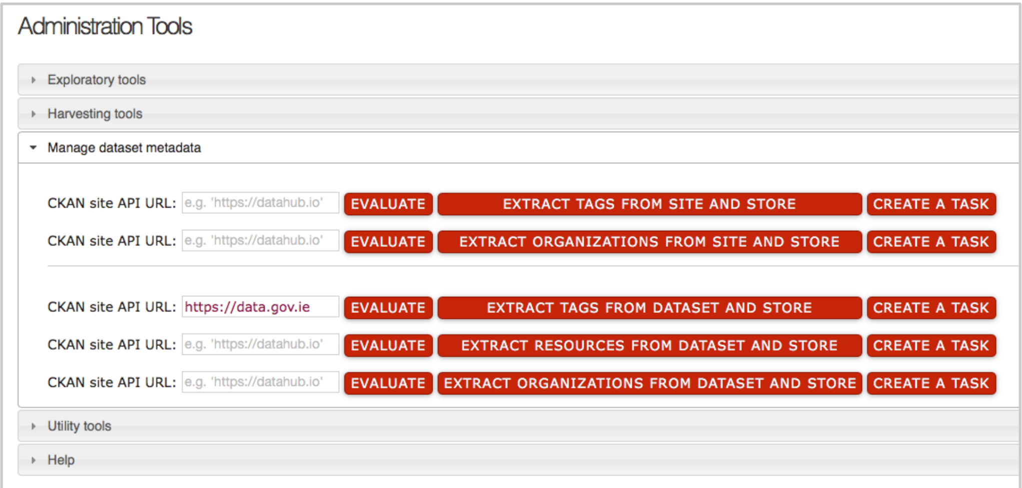



Figure 36 - Admin menu: The management of metadata.

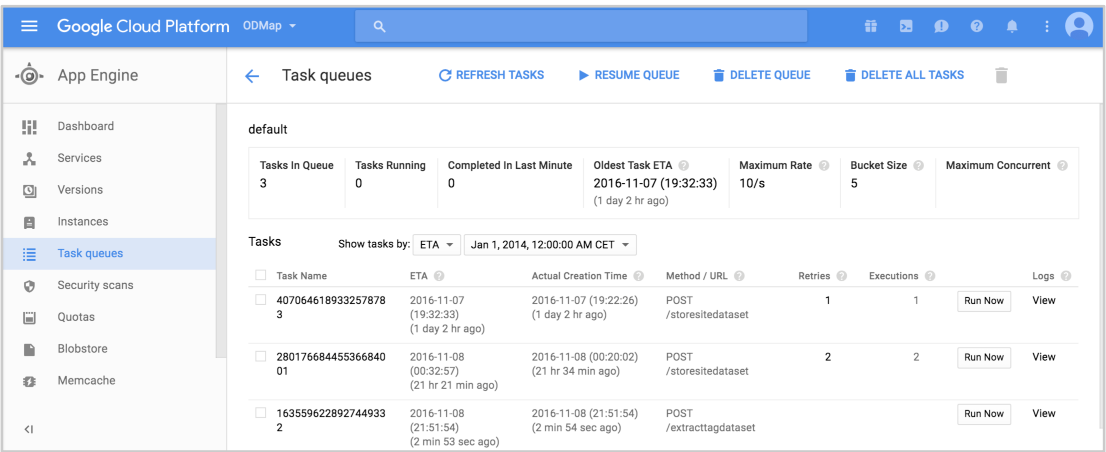



Figure 37 - App Engine Task queues console.

The Table 13 summarizes the main functions of the application in relation to the role of user. Users can be anonymous or registered. In the second case, the user is enabled to perform there are extra functions such as storing the metadata of the dataset of interest in its own repository.

| Function | Anonymous User | Registered User | Admin User |
|---|---|---|---|
| **Open data sites** | | | |
| Search site | ✓ | ✓ | ✓ |
| Browse site repository | ✓ | ✓ | ✓ |
| Manage my data sites [add, remove] | | ✓ | ✓ |
| Manage data sites [harvesting, add, remove, georef, other] | | | ✓ |
| | | | |
| **Datasets** | | | |
| Search dataset | ✓ | ✓ | ✓ |
| Browse dataset metadata repository | ✓ | ✓ | ✓ |
| Go to original dataset | ✓ | ✓ | ✓ |
| Manage my datasets [add, remove] | | ✓ | ✓ |
| Manage datasets [harvesting, add, remove, georef, other] | | | ✓ |
| | | | |
| **Resources** | | | |
| Search resource | ✓ | ✓ | ✓ |
| Browse resource metadata repository | ✓ | ✓ | ✓ |
| Go to original dataset containing a resource | ✓ | ✓ | ✓ |
| Manage my resources [add, remove] | | ✓ | ✓ |
| Manage datasets [harvesting, add, remove, georef, other] | | | ✓ |

Table 13 - ODMap main functions.

The deployment of ODMap in a cloud platform contributes to alleviate the most common technical problems and severely curtails issues associated with scalability and performance, especially when search expands across multiple portals.

# 5 The second case study: NESSIE - A Network-based Environment Supporting Spatial Information's Exploration.

NESSIE is a cloud-based application that supports citizens and stakeholders in evaluating real estate offers through the visualization of data about a building including location, socio-economical data and related OGD such as cadastral data, urban plans, etc. [103]. Usually, the evaluator is interested in knowing the selling price of similar areas in the same region, looks for real estate offers in the Web, asks real estate agencies, consults urban plans, etc. Unlike what happens in other domains such as stock exchange, analytical models and reference values about the real estate market offer a low-level support because the evaluation strongly depends on both the geographical context and commercial aspects. For example, if a building is located near to a highway, this could be appreciated by an enterprise, which deals his products on a daily basis. Conversely, this aspect could be negative for a citizen who aims to rent an apartment. As such the context of data involves ascribe meaning to both data and spatial information [104] [105] [106].

Figure 38 shows the architecture of NESSIE. According to the proposed framework, the datastore contains the catalogue of data and the procedural component consists in services that implement a set of function also called Spatial Decision Support System (SDSS). Data Management Services capture and pre-process data from several web sources including Italian real estate agencies (immobiliare.it, subito.it) and sites related to real estate market in USA (trulia.com and zillow.com), municipalities and government agencies [107][108].

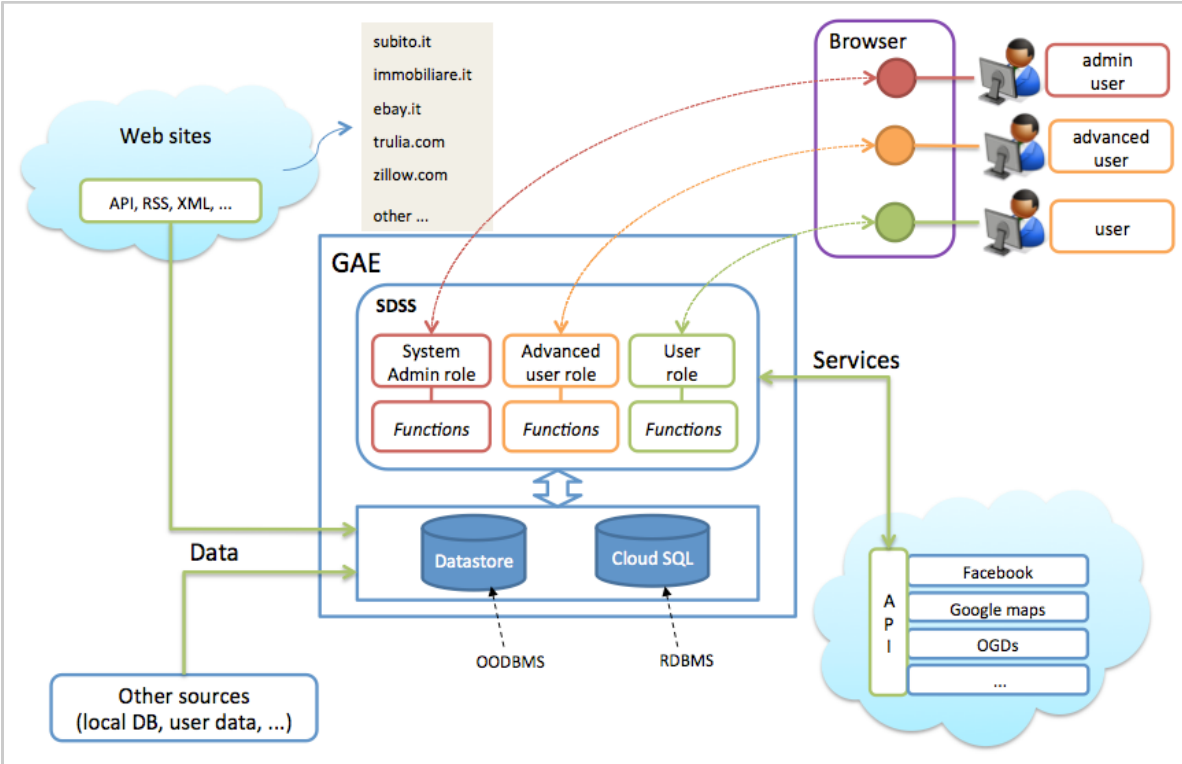



Figure 38 - The Architecture of NESSIE.

To extract data, the user details the geographic zone (region, state, city, location) and the kind of real estate property (apartment, area, building, etc.). Then, the application subscribes an RSS service and captures data from the selected real estate site. As data may have different shortcomings (i.e. they are often partially structured or miss important information such as the geo-tag of a property) they are initially staged in a buffering area, validated and then stored in the Datastore by a specialized service that parses data and automatically feeds the Catalogue.

Figure 39 shows an example data extracted from www.trulia.com. Figure 40 show the logical flow of the data validation.

```
---------------------------------------------------------------------------------------
2064 Jackson St, San Francisco, CA 94109, $7,495,000 3 beds, 2.5 baths
Mon, 14 Nov 2016 08:00:00 GMT
https://www.trulia.com/property/3255324042-2064-Jackson-St-San-Francisco-CA-94109
---------------------------------------------------------------------------------------
52 Palm Ave, San Francisco, CA 94118, $2,395,000 3 beds, 2 baths
Mon, 14 Nov 2016 08:00:00 GMT
https://www.trulia.com/property/3255325473-52-Palm-Ave-San-Francisco-CA-94118
---------------------------------------------------------------------------------------
2136 Beach St, San Francisco, CA 94123, $2,595,000 3 beds, 3 baths
Mon, 14 Nov 2016 08:00:00 GMT
https://www.trulia.com/property/3255313782-2136-Beach-St-San-Francisco-CA-94123
---------------------------------------------------------------------------------------
433 Buena Vista Ave, San Francisco, CA 94117, $649,000 1 bed, 1 bath
Mon, 14 Nov 2016 08:00:00 GMT
https://www.trulia.com/property/3255313920-433-Buena-Vista-Ave-San-Francisco-CA-94117
---------------------------------------------------------------------------------------
```

**Figure 39 - Data extracted from www.trulia.com**

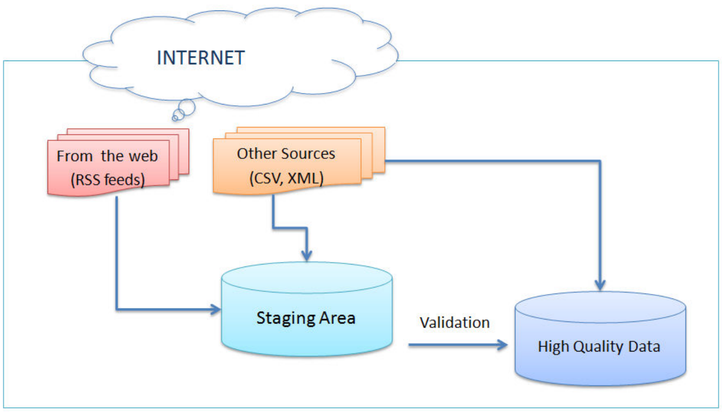



**Figure 40 - The logical flow of the data validation process.**

Table 14 shows a list of implemented task-oriented services.

| Name | Functionality |
|---|---|
| *HouseList1* | Management of the final version of real estate data |
| *Houselist2* | Management of current real estate deals |
| *Housemap* | Visualization of the real estate entities on a map |
| *Housesetup* | Insertion of a new real estate entity |
| *Housegeoref* | Auto-geo-referentiation of a real estate entity from its address |
| *Housestage* | Validation of captured real estate entities |
| *NewURL* | Insertion of a new URL of a web site for data feeding |
| *URL* | List and selection of available web sites for data feeding |

**Table 14 - List of implemented task-oriented services.**

Special zones (for example cadastral zones or detailed user areas) can be imported from external files. Users are enabled to visualize single maps, define and store geographical zones highlighted with different colours, aggregate different zones in a single zone, namely a macro-zone.

Figure 41 shows how NESSIE presents information about the macro-zones defined by an urban plan for “San Francisco - Downtown” (USA).

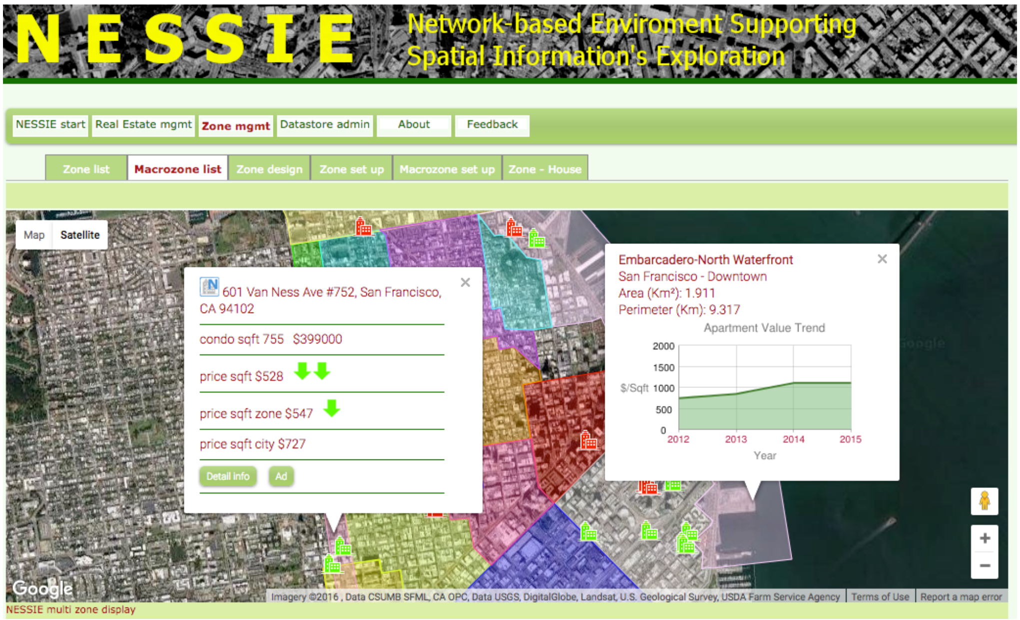



Figure 41 - NESSIE: a macro-zone in San Francisco (USA).

The spot on the right gives information about a specific zone (“Embarcadero-North Waterfront”) in which the user is interested. Specifically, the spot shows the area and the perimeter of the macro-zone and details the trend of the apartment value (i.e. the price per square feet) during the last four years. Real estate properties are highlighted with different coloured graphic symbols. When the user clicks on a symbol, it produces the spot on the left-side of Figure 41 which shows the address of the property, its area and cost as well statistical information about the average price of real estate properties in the zone where the property is located and in the metropolitan area of San Francisco.

Figure 42 shows and additional example of the macro-zones defined by the urban plan of Cagliari (Italy) and its hinterland.

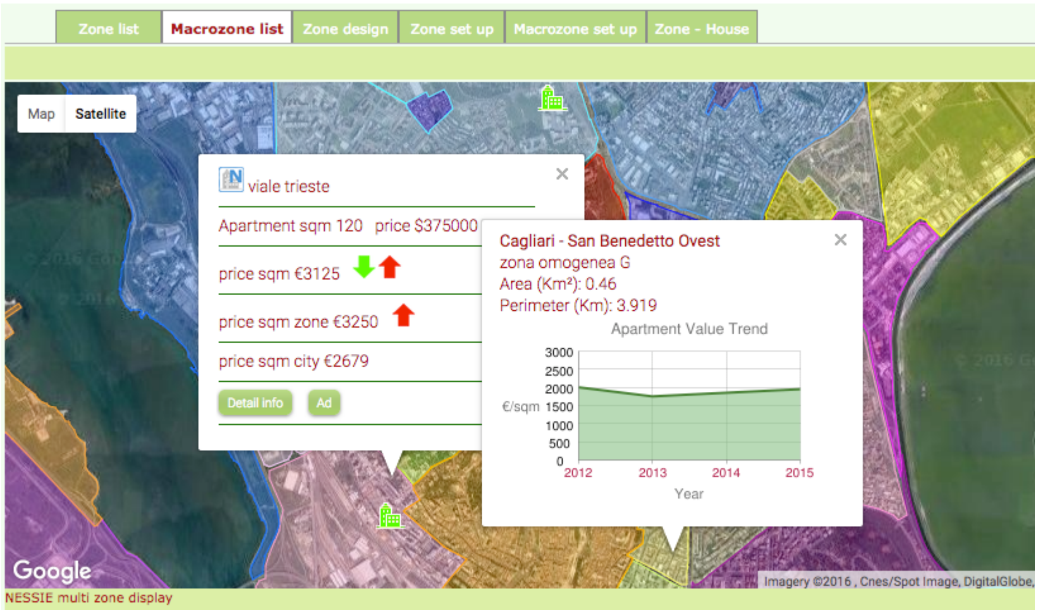



Figure 42 - Data support within a macro-zone.

Selecting a property in a pre-fixed zone or macro-zone may eventually lead to the proposal of offers whose prices are compared to the price in the same zone. Usually, these zones are cadastral areas as defined by the urban plans. It is not unlikely that the user wants to acquire information on a new zone that crosses many cadastral zones. He can do it by defining interactively a customized zone of arbitrary form, namely a "dynamic spatial context".

Figure 43 shows an example of dynamic spatial context. Here the user draws interactively a polygonal bounding that is shaped like a star. Visualized information is only about this context: the green houses represent properties whose cost is lower than the medium cost of the similar properties offered in the polygonal bounding while the red houses represent properties whose cost is higher.

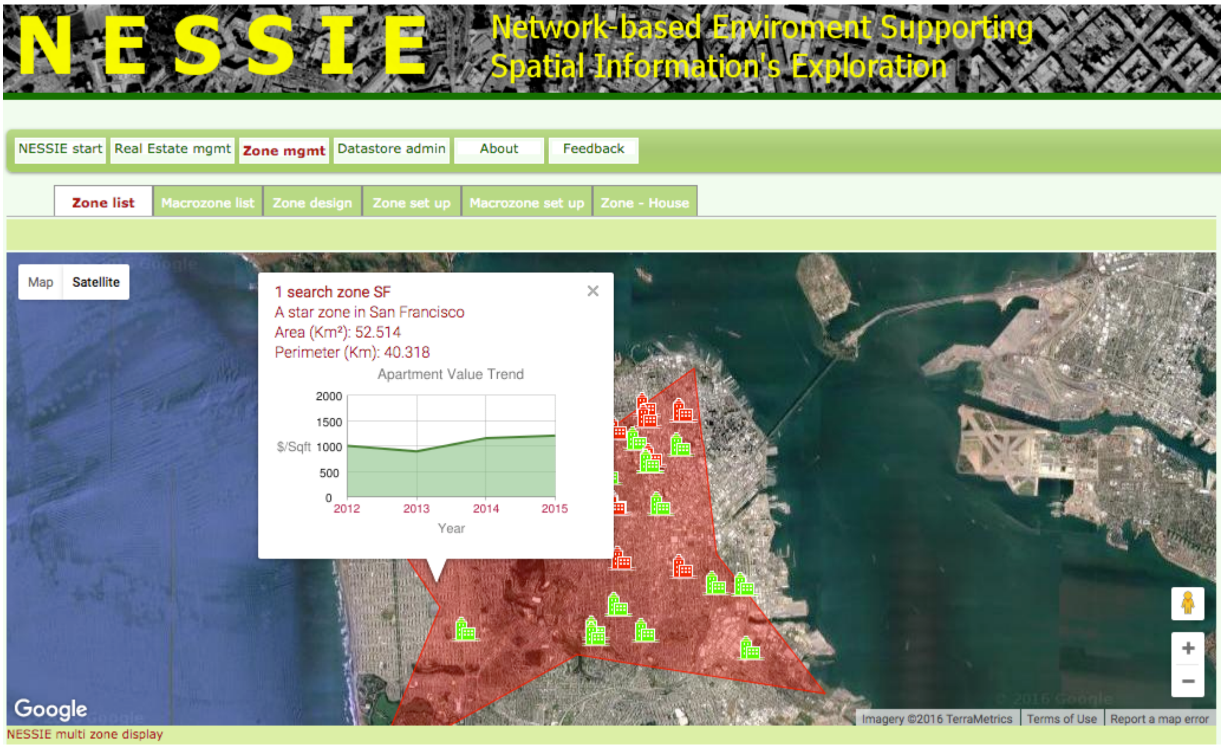



Figure 43 - NESSIE: Example of dynamic spatial context.

By means of Facebook, users are allowed to store permanently and share maps and spatial contexts (i.e. the maps, the polygonal bounding and the overlaid reference information) with their friends, as we will explain in what follows.

Additionally, Figure 44 lists the zones the user shares with his MyNESSIE friends.

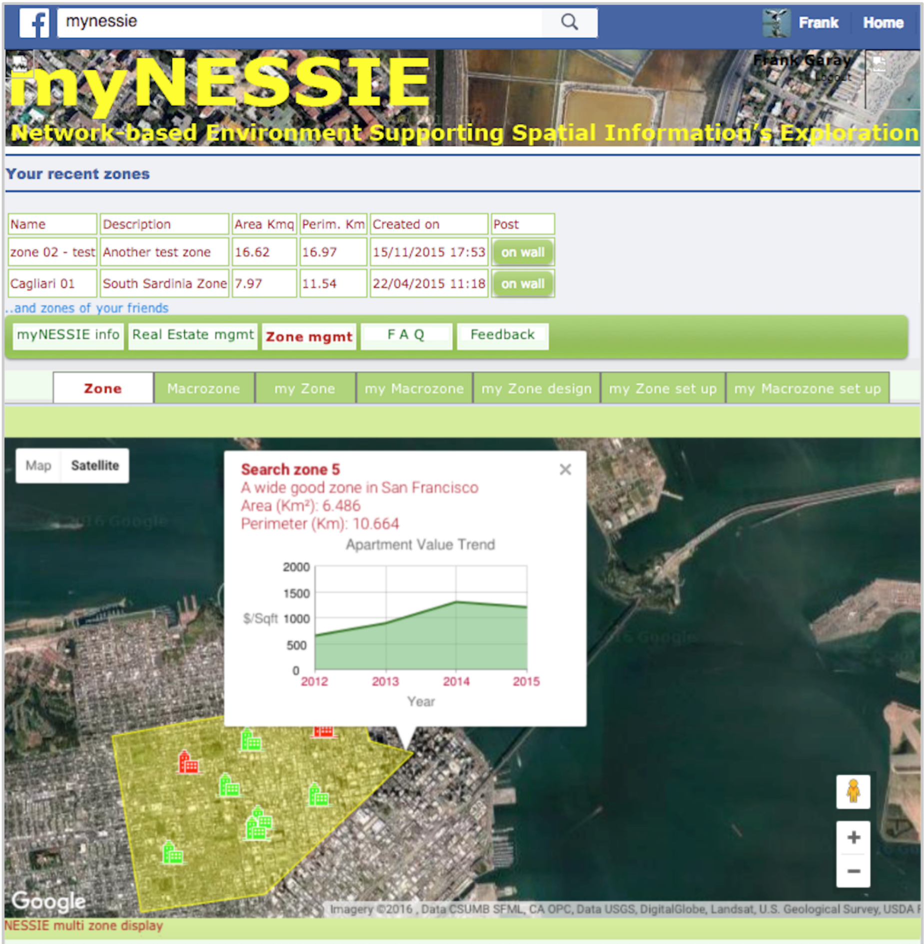



Figure 44 - MyNESSIE: example of dynamic spatial context.

We note that these zones are not necessarily in San Francisco, but in different geographical areas, which the user and his friends are interested in. Collaboration happens by means of Facebook whose users are enabled to create lists for staying in touch with other users and organize their friends as they like. To access MyNESSIE, a Facebook user is required to make a subscription that stores, into an object of the kind "User", data about the subscriber and the list of his friends. Subscribed users are enabled to share maps and objects only with their friends who, in turn, subscribed MyNESSIE. As Figure 45 shows, they can see more of them in their news feed and get notified each time new MyNESSIE objects are posted.

Different kinds of maps are available. Figure 46 shows an additional example of user-defined zone on a topological map.

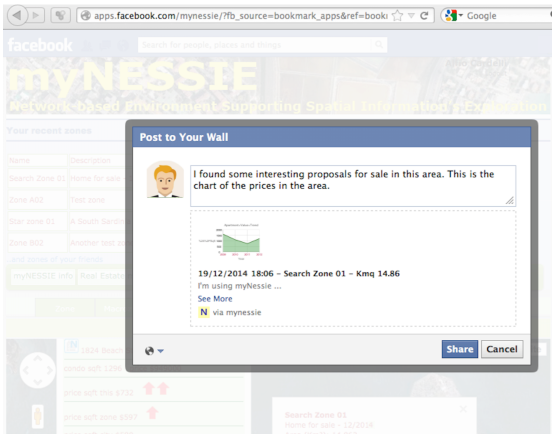



Figure 45 - MyNESSIE: user notification message.

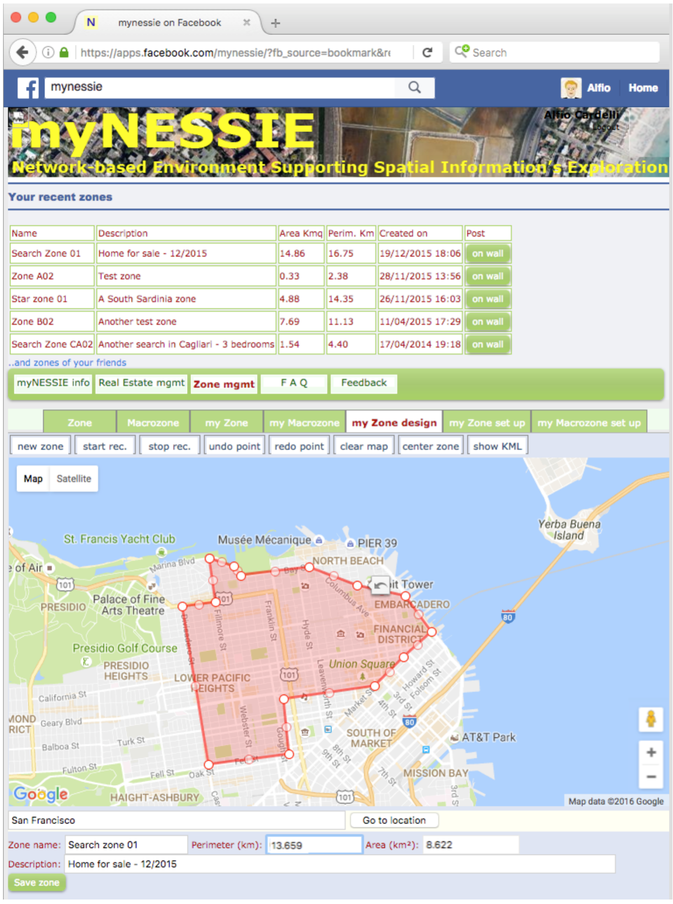



Figure 46 - MyNESSIE: additional example of user-defined zone.

# 6 Conclusions

The open data movement as well as recent advances in ICT technologies are promoting the introduction of new computational paradigms that will have great influence on providing transparency in government as well as innovative, and efficient information and services management solutions for publishing and opening government data.

This thesis aimed to highlight the importance of such computational paradigms especially while acquiring information that consider user preferences or making data reusable by applications that affect the interest of various stakeholders and impact value creation in terms of creating new ways of using information.

Giving attention to state-of-the-art publishing platforms and Cloud Computing, new paradigms should concentrate not only on intelligent searching and integrating data gathered from several sources but should enable frameworks that consider data structures and architectural solutions in which such data may be made relevant to citizens, application developers and ultimately to the information society.

In this thesis, the proposed framework is based on cloud computing and considers architectural aspects and data management technologies which are of great significance both for scientific and applicative points of view.

In particular, the framework relies on APIs for capturing data and loosely structures contextual information coming from diverse resources using a flexible data and database management offered as cloud service. It aims to be a reference point for the development of applications that derive new insights from data and allow to combine government data sources and information gathered from other web resources.

The proposed framework presents the following main features:

- is a general framework based on currently available cloud environments;
- deals with general scenarios for publishing OGD;
- aims to be a reference framework for developing OGD-based apps on top of OGD portals and other web resources;
- supports the deployment of cloud-native applications without constraints about the type of application;
- relies on a metadata catalogue so that the resource provision takes place when applications are deployed.

These advantages should encourage developers and stakeholders to create value added applications that could positively impact the quality of their work. Using traditional architectures, the capacity of creating such applications requires accurate evaluations of computing needs and could not be adjusted easily, contributing to inefficient utilization of OGD resources. Indeed, the cloud environment can be easily altered to accommodate changes in web resources as it enables the development of applications featured by load scalability (the cloud environment expands and contracts its resource pool to accommodate heavier or lighter loads) and geographic scalability (the cloud environment maintains usefulness and usability regardless of how far apart its users or resources are).

Cloud applications that fit at the best the proposed framework were also presented and discussed from an experimental point of view. Two case studies explore the prospective of implementing flexible and easy-to-use cloud applications that enable users to search OGD in an interactive way, change and refine their preferences, perform personal evaluations in a real-time manner and share information by means of a social network.

This thesis aims to trace a road for the deployment OGD-based apps in cloud environment: it tries to identify the nature of the technology we need in order to promote the development of value added OGD application by government agencies, developers and stakeholders.

The proposed framework is in the early stage of its validation. Additional studies will be carried out in the next future to intensively test the developed applications, and considerable effort will be invested to evaluate the reliability and scalability of the proposed solution as the catalogue grows, in order to allow a comparison with traditional web applications.

Most important, we have tried to identify the nature of the technology we need in order to promote the development of collaborative and specialized computational solutions to support decisions in collaborative manner involving customers, professional and domain experts.

# 7 References


[1] Obama, B.: Transparency and Open Government - Memorandum for the Heads of Executive Departments and Agencies, 2009.01.21. Retrieved May, 2015. *https://www.whitehouse.gov/open/documents/open-government-directive*.

[2] Colpaert, P., Dimou, A., Sande, M.V., Breuer, J., Van Compernolle, M., Mannens, E., Mechant, P., Van de Walle, R.: A three-level data publishing portal. Athens: European Data Forum (2014). Retrieved May 2015. *http://2014.dataforum.eu/sites/default/files/pdf/edf2014_submission_43.pdf.*

[3] Grimmelikhuijsen, S.G., Porumbescu, G., Hong, B., Im, T.: The effect of transparency on trust in government: a cross-national comparative experiment. Public Adm. Rev. 73(4), pp. 575–586, 2013.

[4] Meijer, A.: Understanding the complex dynamics of transparency. Public Adm. Rev. 73(3), pp. 429–439, 2013.

[5] Weerakkody, V., Janssen, M., Dwivedi, Y.K.: Transformational change and business process reengineering (BPR): lessons from the British and Dutch public sector. Gov. Inf. Q. 28, pp. 320–328, 2011.

[6] Linders, D., Wilson, S.C.: What is open government? One year after the directive. In: 12th Annual International Conference on Digital Government Research (Dg.o 2011), pp. 262–271, 2011.

[7] Grimmelikhuijsen, S., Porumbescu, G., Hong, B., Im, T.: The effect of transparency on trust in government: a cross-national comparative experiment. Public Adm. Rev. 73(4), pp. 575–586, 2013.

[8] Martin, S., Foulonneau, M., Turki, S.: 1-5 Stars: Metadata on the Openness Level of Open Data Sets in Europe. In: Garoufallou, E., Greenberg, J. (eds.) MTSR 2013. CCIS, vol. 390, pp. 234–245. Springer, 2013.

[9] Dulong de Rosnay, M., Janssen, K.: Legal and Institutional Challenges for Opening Data across Public Sectors: Towards Common Policy Solutions. Journal of theoretical and applied electronic commerce research, vol. 9, issue 3, pp. 1-14, (09 2014).

[10] Yu, H., Robinson, D.: The new ambiguity of ‘Open government’. UCLA Law Rev. Discl. 59, pp. 178–208, 2012.

[11] Dublin Core Metadata Initiative. DCMI Specification page. Retrieved December, 2015. *http://dublincore.org/specifications*.

[12] Conradie, P., Choenni, S.: Exploring process barriers to release public sector information in local government. In: Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance. pp. 5–13. ICEGOV '12, ACM, New York, NY, USA, 2012, *http://doi.acm.org/10.1145/2463728.2463731.*

[13] Zuiderwijk, A., Janssen, M.: Barriers and development directions for the publication and usage of open data: A socio- technical view. In: Gascó-Hernández, M. (ed.) Open Government, Public Administration and Information Technology, vol. 4, pp. 115–135. Springer New York, 2014, *http://dx.doi.org/10.1007/978-1-4614-9563-5_8.*

[14] Google App Engine. Retrieved May, 2014. *https://cloud.google.com/appengine.*

[15] Electronic Frontier Foundation: History of FOIA. Retrieved May, 2015. *https://www.eff.org/issues/transparency/history-of-foia.*

[16] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. Retrieved March, 2015. *http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:02003L0098-20130717*

[17] The annotated 8 Principles of Open Government Data. Home page. Retrieved March, 2015. *https://opengovdata.org*.

[18] Data.gov Home Page. Retrieved December, 2015. *https://www.data.gov*.

[19] Cameron, D.: Letter to Government departments on opening up data (2010). Retrieved March, 2015. *https://www.gov.uk/government/news/letter-to-government-departments-on-opening-up-data.*

[20] Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information. Retrieved March, 2015. *http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32013L0037*.

[21] About Open Government Partnership. Retrieved December, 2015. *http://www.opengovpartnership.org/about.*

[22] Open Data Charter History. Retrieved December, 2015. *http://opendatacharter.net/history*.

[23] Janssen, K.: The influence of the PSI directive on open government data: An overview of recent developments. Government Information Quarterly, 28(4), pp. 446-456, 2011.

[24] European Commission: Guidelines on recommended standard licences, datasets and charging for the reuse of documents. Commission Notice, Official Journal of the European Union C240/1, 24 July 2014.

[25] Berners-Lee, T.: Linked Data, 2006. Retrieved May, 2014. *http://www.w3.org/DesignIssues/LinkedData.html*.

[26] Attard, J., Orlandi, F., Scerri, S., Auer, S.: A systematic review of open government data initiatives. Government Information Quarterly, Volume 32, Issue 4, pp. 399–418, 2015.

[27] ODI Open Data Institute. Retrieved May, 2016. *http://theodi.org/about*.

[28] Sunligth Foundation Home Page. Retrieved May, 2016. *https://sunlightfoundation.com*.

[29] Sunligth Foundation. Open Data Policy Guidelines. Retrieved May, 2016. *https://sunlightfoundation.com/opendataguidelines*.

[30] Europe's Public Data Home Page. Retrieved May, 2016. *http://publicdata.eu*.

[31] European Data Portal Home Page. Retrieved September, 2016. *http://www.europeandataportal.eu*.

[32] OKFN. Open Data Census page. Retrieved September, 2016. *http://census.okfn.org/en/latest*.

[33] Pew Research Center: Americans' Views on Open Government Data, 2015.04.20. Retrieved September, 2015. *http://www.pewinternet.org/2015/04/21/open-government-data/pi_2015-04-21_open-gov-data_06/.*

[34] Mintz, D.: Government 2.0 - Fact or Fiction? The Public Manager, Winter 2007/2008, 36(4) pp. 21-24.

[35] Bonsón, E., Torres, L., Royo, S., Flores, F.: Local e-government 2.0: Social media and corporate transparency in municipalities. Government Information Quarterly, 29(2), pp.123–132, 2012.

[36] Davies, T., Frank, M.: There's no such thing as raw data: Exploring the socio-technical life of a government dataset. Proceedings of the 5th Annual ACM Web Science Conference, pp. 75–78. New York, NY, USA: WebSci '13, ACM, 2013.

[37] Hendler, J., Holm, J., Musialek, C., Thomas, G.: Us government linked open data: Semantic.data.gov. IEEE Intelligent Systems, 27(3), pp. 25–31, 2012.

[38] Van der Veer, H., Wiles, A.: Achieving Technical Interoperability - the ETSI Approach. ETSI White Paper No.3, 3rd edition, April 2008. Retrieved May, 2015. *https://portal.etsi.org/CTI/Downloads/ETSIApproach/IOP whitepaper Edition 3 final.pdf.*

[39] Martin, S., Foulonneau, M., Turki, S.: 1-5 stars: Metadata on the openness level of open datasets in Europe. In: Garoufallou, E., Greenberg, J. (eds.) MTSR 2013. Communications in Computer and Information Science, vol. 390, pp. 234-245, Springer, 2013.

[40] Marienfeld, F., Schieferdecker, I., Lapi, E., Tcholtchev, N.: Metadata aggregation at Govdata.de - An Experience Report. In: Proceedings of the 9th International Symposium on Open Collaboration, art. 21, pp. 1-5. WikiSym '13, ACM, New York, NY, USA, 2013. *http://dx.doi.org/10.1145/2491055.2491077*.

[41] Open Knowledge International: Global open data index – Place overview. Retrieved September, 2015. *http://index.okfn.org/place*.

[42] Hollyer, J. R., Rosendorff, B. P., Vreeland J. R.: Measuring Transparency. Political Analysis (2014) 22 (4), pp. 413-434. *https://doi.org/10.1093/pan/mpu001*.

[43] HRV Transparency Project. Retrieved May, 2016. *http://hrvtransparency.org*.

[44] Kosack, S., Fung, A.: Does Transparency Improve Governance? In: Annual Review of Political Science, Vol.17, pp. 65-87, 2014. DOI: 10.1146/annurev-polisci-032210-144356.

[45] Open Data Barometer Home Page. Retrieved May, 2016. *http://opendatabarometer.org*.

[46] World Wide Web Foundation Home Page. Retrieved May, 2016. *http://webfoundation.org*.

[47] Open Data Barometer: ODB Methodology v1.0 28th April 2015. Retrieved May, 2016. *http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-Methodology.pdf*.

[48] Zuiderwijk, A., Janssen, M.: The negative effects of open government data - Investigating the dark side of open data. Proceedings of the 15th Annual International Conference on Digital Government Research (dg.o '14), pp. 147–152. New York, NY, USA, ACM, 2014. *http://dx.doi.org/10.1145/2612733.2612761*.

[49] Zuiderwijk, A., Janssen, M.: Barriers and Development Directions for the Publication and Usage of Open Data: A Socio-Technical View. In M. Gascó-Hernández (Ed.), Open Government, Public Administration and Information Technology, vol. 4, pp. 115–135, New York, Springer, 2014. *http://dx.doi.org/10.1007/978-1-4614-9563-5_8*.

[50] Dessì, N., Garau, G., Pes, B.: Web 2.0 Technologies Empowering Spatial Decision Making. Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2012 IEEE 21st International Workshop on, Toulouse, 2012, pp. 462-467. doi: 10.1109/WETICE.2012.31

[51] Sivarajah, U., Irani, Z., Weerakkody, V.: Evaluating the use and impact of Web 2.0 technologies in local government. Government Information Quarterly, 32, pp. 473–487, (2015). *http://dx.doi.org/10.1016/j.giq.2015.06.004*.

[52] Mauher, M., Smokvina, V.: Digital to Intelligent Local Government Transition Framework. In: Proceedings of the 29th International Convention of MIPRO, Opatija, Croatia, May 22-26, 2006.

[53] Bischof, S., Karapantelakis, A., Nechifor, C.-S., Sheth, A., Mileo, A., Barnaghi, P.: Semantic Modelling of Smart City Data. W3C Workshop on the Web of Things Enablers and services for an open Web of Devices, 2014.

[54] Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, pp.1-22, 2009.

[55] Consoli, S., Gangemi, A., Nuzzolese, A., Peroni, S., Presutti, V., Reforgiato, D., Spampinato, D.: Geolinked Open Data for the Municipality of Catania. In: Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14), 2014.

[56] Alexopoulos, C., Zuiderwijk, A., Charapabidis, Y., Loukis, E., Janssen, M.: Designing a Second Generation of Open Data Platforms: Integrating Open Data and Social Media. Electronic Government, LNCS 8653, pp. 230–241, Springer, 2014.

[57] Duval, A., Brasse, V.: How to ensure the economic viability of an open data platform. Procedia Computer Science, 33, pp. 179–182, 2014.

[58] Baldwin, C. Y., Woodard, C. J.: The architecture of platforms: A unified view. Harvard Business School, 2009.

[59] Tiwana, A., Konsynski, B., Bush, A.: Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. Information Systems Research, 21(4), pp. 675–687, 2010. *http://doi.org/10.1287/isre.1100.0323*.

[60] Data Catalog Vocabulary (DCAT). W3C Recommendation 16 January 2014. Retrieved April, 2014. *https://www.w3.org/TR/vocab-dcat*.

[61] World Bank: Technical assessment of open data platforms for national statistical organisations. Washington, DC: World Bank Group, 2014. Retrieved May, 2015. *http://documents.worldbank.org/curated/en/744241468334210686/Technical-assessment-of-open-data-platforms-for-national-statistical-organisations*.

[62] ROUTE-TO-PA Project: Raising Open and User-friendly Transparency-Enabling Technologies for Public Administrations, Project number 645860, H2020-INSO-2014. *http://routetopa.eu/*.

[63] ROUTE-TO-PA Project: D2.1 State-of-the-art Report and Evaluation of Existing Open Data Platforms (Version 1.1 – Revised Version – 19/05/2016). Retrieved September, 2015. *http://routetopa.eu/wp-content/uploads/2015/06/D2.1-State-of-the-art_Report_and_Evaluation_of_Existing_Open_Data_Platforms-v1.1.pdf*.

[64] Maali, F., Cyganiak, R., Peristeras, V.: Enabling Interoperability of Government Data Catalogues. In: Electronic Government, Proceedings of the IFIP WG 8.5 International Conference, EGOV 2010, Lausanne, Switzerland, August 29 - September 2, 2010, LNCS 6228, pp. 339-350, Springer, 2010.

[65] Data.gov.uk Home Page. Retrieved December, 2015. *https://data.gov.uk*.

[66] CKAN The open source data portal software. Retrieved December, 2015. *http://ckan.org*.

[67] CKAN Association – Managing and Overseeing CKAN. Retrieved December, 2015. *http://ckan.org/about/association*.

[68] CKAN instances around the world. Retrieved December, 2015. *http://ckan.org/instances*.

[69] DKAN The open source open data platform. Retrieved March, 2016. *http://www.nucivic.com/dkan*.

[70] CKAN Project on GitHub. Retrieved December, 2015. *https://github.com/ckan/ckan*.

[71] Hoppin, A., Byrnes, A., Couch, A.: Open-Source Open Data Platforms - Circa 2013. Retrieved March, 2016. *http://www.w3.org/egov/wiki/images/f/f1/W3C_OpenSource_OpenData.pdf.*

[72] CartoDB Home Page. Retrieved March, 2015. *http://cartodb.com*.

[73] Nikolaos Loutas, Anastasia Varitimou, Vassilios Peristeras, V.: Unraveling the mystery of Open Government Data Apps. W3C workshop on USING OPEN DATA policy modeling, citizen empowerment, data journalism. Workshop paper, 2012. Retrieved March, 2015. *http://www.w3.org/2012/06/pmod/pmod2012_submission_11.pdf.*

[74] Lee, M., Almirall, E., Wareham, J.: Open Data and Civic Apps: First-Generation Failures, Second-Generation Improvements. Communications of the ACM, 59(1), pp. 82-89, 2016.

[75] DontEat Home Page. Retrieved May, 2015. *http://www.donteat.at*.

[76] ABCEats iTunes preview. Retrieved May, 2015. *https://itunes.apple.com/us/app/abceats/id502867547?mt=8*.

[77] City of Edmonton Apps page. Retrieved May, 2015. *https://www.edmonton.ca/programs_services/city-of-edmonton-apps.aspx*.

[78] Street Bump Home Page. Retrieved May, 2015. *http://www.streetbump.org*.

[79] FixMyStreet Platform. Retrieved May, 2015. *http://fixmystreet.org*.

[80] Roadify Transit Home Page. Retrieved May, 2015. *http://www.roadify.com*.

[81] OpenSpending Home Page. Retrieved June, 2015. *https://openspending.org*.

[82] Where Does My Money Go? Retrieved June, 2015. *http://app.wheredoesmymoneygo.org*.

[83] Rodríguez-Bolívar, M.P., (ed.): Measuring E-government Efficiency - The Opinions of Public Administrators and Other. Public Administration and Information Technology. Vol. 5, Springer, 2014.

[84] The White House: The Obama's Administration Commitment to Open Government: a status report. 2011. Retrieved February, 2015. *http://www.whitehouse.gov/sites/default/files/opengov_report.pdf*.

[85] Lin, G., Fu, D., Zhu, J., Dasmalchi, G.: Cloud Computing: IT as a Service. IT Professional 11(2), pp. 10-13, 2009.

[86] Taylor, R.C.: An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. BMC Bioinformatics 11(Suppl 12):S1, 2010. *http://dx.doi.org/10.1186/1471-2105-11-S12-S1*.

[87] Chae, H., Jung, I., Lee H., Marru S., Lee S. W., Kim S.: Bio and health informatics meets cloud: BioVLab as an example. Health Information Science and Systems, pp. 1-6, 2013. *http://dx.doi.org/10.1186/2047-2501-1-6*.

[88] Matz, K.: Designing Usable Apps - An agile approach to User Experience design. Winchelsea Press, 2013.

[89] Dessì, N., Fugini, M. G., Garau, G., Pes, B.: Architectural and Security Aspects in Innovative Decisional Supports. In: Proceedings of ITAIS'13 Conference, Milano, Dec. 13, 2013.

[90] Grossman, R.L., White, K.P.: A vision for a biomedical cloud. Journal of Internal Medicine 271(2), pp. 122-130, 2012.

[91] Iemma, R., Morando, F., Osella, M.: Breaking Public Administrations' Data Silos: The Case of Open-DAI, and a Comparison between Open Data Platforms. JeDEM, 6(2), pp. 112-122, 2014. Retrieved March, 2015. *http://www.jedem.org/index.php/jedem/article/view/304/27*.

[92] Antikainen, M., Mäkipää, M., & Ahonen, M.: Motivating and supporting collaboration in open innovation. European Journal of Innovation Management, 13(1), pp. 100-119, 2010. *http://doi.org/10.1108/14601061011013258*.

[93] Open Government Data Initiative v6. Retrieved January, 2015. *https://github.com/openlab/OGDI-DataLab*.

[94] About OGDI DataLab. Retrieved January, 2015. *http://ogdifrance.cloudapp.net/Home/About*.

[95] Jing, H., Haihong, E., Guan, L., Jian, D.: Survey on NoSQL database. In 6th International Conference on Pervasive Computing and Applications (ICPCA), pp. 363-366, 2011.

[96] Dessì, N., Garau, G., Reforgiato R., D., Pes, B.: Increasing open government data transparency with spatial dimension. 2016 IEEE 25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Paris, 2016, pp. 247-249. doi: 10.1109/WETICE.2016.61.

[97] SIOPE Sistema Informativo delle Operazioni degli Enti Publici. Retrieved December, 2014. *https://www.siope.it.*

[98] Stonebraker, M.: SQL databases v. NoSQL databases. Communications of ACM, 53(4), pp. 10-11, April 2010.

[99] CKAN API Guide. Retrieved June, 2015. *http://docs.ckan.org/en/latest/api*.

[100] OpenLayers 3 Home page. Retrieved July, 2015. *https://openlayers.org*.

[101] Leaflet Home page. Retrieved July, 2015. *http://leafletjs.com*.

[102] ODMap - Open Data Mapping. Available on *http://odmap-unica.appspot.com*.

[103] NESSIE - Network-based Enviroment Supporting Spatial Information's Exploration. Available on *http://nessie-hr.appspot.com*.

[104] Rinner, C., Keßler, C., Andrulis, S.: The Use of Web 2.0 Concepts to Support Deliberation in Spatial Decision-Making. Computers, Environment and Urban System, 32(5), pp. 386-395, 2008.

[105] Hu, L., Yue, P., Zhou, H.: Geoprocessing in Google Cloud Computing: case studies. 2012 First International Conference on Agro-Geoinformatics (Agro-Geoinformatics). pp. 1-6, Shangai, 2012. doi: 10.1109/Agro-Geoinformatics.2012.6311653

[106] Steiniger, S., Poorazizi, M. E., Bliss-Taylor, C. A. M., Mohammadi, E., Hunter, A. J. S.: PlanYourPlace: Merging Social Networks and Participatory GIS for Participatory Planning. Proceedings of FIG Working Week 2012. Rome, Italy, 6-10 May 2012.

[107] Argiolas, M., Dessì, N., Garau, G., Pes, B.: Web 2.0 services for collaborative spatial decision making. International Journal of Collaborative Enterprise. Vol. 3(2), pp. 189-207, Genève, 2013. doi: 10.1504/IJCENT.2013.053295.

[108] Dessì, N., Garau, G.: A Case Study for a Collaborative Business Environment in Real Estate. New Trends in Databases and Information Systems. Advances in Intelligent Systems and Computing. Vol. 241, pp. 351-360, Springer, 2014.